

Diseño estadístico de experimentos

2.ª edición

Abel Díaz

Ciencia y Tecnología

Abel Díaz Cadavid

Diseño estadístico de experimentos

2.ª edición

Colección *Ciencia y Tecnología*

ISBN: 978-958-714-264-8

Primera edición: junio de 1999
Segunda edición: diciembre de 2009
Fotografía y diseño de cubierta: Anastasia Garzón Vidal, Imprenta Universidad de Antioquia
Diagramación: Luz Elena Ochoa Vélez
Impresión y terminación: Imprenta Universidad de Antioquia
Coordinación editorial: Gonzalo Montoya Velásquez

Impreso y hecho en Colombia / Printed and made in Colombia

Editorial Universidad de Antioquia
Teléfono: (574) 219 50 10. Telefax: (574) 219 50 12
Correo electrónico: editorial@udea.edu.co
Sitio web: http://editorial.udea.edu.co
Apartado 1226. Medellín. Colombia

Ficha bibliográfica: véase página iv.

Contenido

Introducción

Convencido de la utilidad y pertinencia de un texto guía para los cursos de diseño de experimentos dictados en la Universidad de Antioquia, y considerando la escasa bibliografía disponible en nuestro medio, me propuse escribir este libro dirigido, en particular, a los estudiantes que reciben un segundo curso en bioestadística y, en general, a estudiantes, investigadores y profesores que requieran diseñar experimentos y analizar datos estadísticamente.

El libro se enfoca hacia la estadística experimental, no desarrolla procedimientos matemáticos y solamente exige, por parte del lector, los conocimientos mínimos de estadística descriptiva e inferencial, equivalentes a los que normalmente se enseñan en un primer curso de estadística. Aunque la mayoría de los ejemplos y de los ejercicios provienen del área de la biología, el estudiante de otras ciencias tendrá suficiente madurez y poca dificultad en visualizar ejemplos similares en sus propias áreas de estudio.

Los procedimientos estadísticos que se desarrollan en el libro pueden efectuarse con una calculadora de bolsillo. Sin embargo, esos mismos procedimientos pueden agilizarse, y la probabilidad de cometer errores aritméticos es mucho menor, si se usa el computador; esta herramienta es casi indispensable cuando se desea analizar gráficamente los datos.

Los ejemplos y ejercicios numéricos que ilustran los procedimientos del análisis estadístico contienen datos reales, no ficticios, y muchos de ellos son los resultados de investigaciones realizadas por personas vinculadas a las universidades de Medellín.

El libro contiene doce capítulos en los cuales se discuten los temas fundamentales del diseño y análisis de experimentos. En los tres primeros se fundamenta el diseño propiamente dicho. Estos son capítulos básicamente conceptuales donde se establecen las relaciones entre la estadística y el proceso investigativo; se exponen los principios metodológicos y filosóficos del diseño y se muestra su conexidad con la teoría de la inferencia estadística.

En los capítulos 4 y 5 se define y aplica la técnica del análisis de varianza, que se adopta como la metodología que se va a emplear en el análisis de datos

experimentales, y se discute el cumplimiento de aquellos supuestos que dan validez al análisis estadístico.

El capítulo 6 se dedica al análisis de las comparaciones múltiples y ofrece amplia gama de metodologías para las comparaciones por pares. En los capítulos siguientes, del 7 al 10, se discuten los diseños experimentales más simples pero también los más comunes en investigaciones cortas y de carácter exploratorio.

El capítulo 11 presenta al estudiante los principios de la regresión y la correlación lineales. Solo se estudia el caso simple o de una variable a causa de la limitación de espacio y del bagaje matemático requerido. El capítulo 12 resume el análisis tradicional de las variables categóricas, como la prueba de homogeneidad de proporciones y las tablas de contingencia.

Deseo expresar mis agradecimientos muy sinceros a los profesores y estudiantes del Departamento de Biología de la Facultad de Ciencias Exactas y Naturales de la Universidad de Antioquia, que me permitieron escudriñar y seleccionar el material bibliográfico para los ejemplos y ejercicios de este libro. Igualmente, me declaro en deuda de gratitud con todos los investigadores y profesionales que directa o indirectamente me permitieron utilizar sus resultados experimentales para conformar un texto más enfrentado a la realidad y para que su lectura fuese más amena.

Debo, de la misma manera, reconocer y agradecer a mi esposa María Cecilia y a mis dos hijas Marcela y Olga Lucía las privaciones y frecuentes sacrificios en la vida familiar hechos con amor, originados en mi afán de cumplir con el compromiso de ofrecer un libro bien elaborado a la comunidad universitaria.

Este libro fue pensado y bosquejado durante los últimos diez años de docencia universitaria del tema, pero su escritura me fue posible gracias al beneficio del Año Sabático concedido por el Honorable Consejo Superior de la Universidad de Antioquia mediante la Resolución Superior 1501 del 11 de diciembre de 1991.

Abel Díaz Cadavid
Febrero de 1993

1

La estadística en la experimentación

1.1 La investigación científica

La estadística y, en particular, el diseño de experimentos, son considerados en la literatura científica como técnicas que deben incorporarse al desarrollo del proceso investigativo. Ubicar la estadística en el contexto de la investigación es asunto de la teoría del conocimiento. Sin pretender entrar en el campo filosófico, se enunciarán algunas ideas relacionadas con la adquisición del conocimiento científico.

La ciencia, según E. Ander Egg (2001), es el conocimiento racional, cierto o probable, obtenido metódicamente, sistematizado y verificable. Según su objeto, hay dos tipos de ciencia: las ciencias formales, las cuales estudian ideas, como la matemática y la lógica; y las ciencias actuales, que estudian hechos, como las ciencias naturales y humanas.

Esta clasificación tiene en cuenta el método por el cual se ponen a prueba los enunciados de cada ciencia. Mientras las ciencias formales usan el proceso deductivo y tratan con entes ideales, las ciencias fácticas necesitan la observación o la experimentación, o de la observación y la experimentación, para confirmar sus conjeturas; es decir, son ciencias empíricas y objetivas.

Por otra parte, según el tipo de interés que se tenga en la búsqueda del conocimiento, la ciencia tiene dos propósitos: incrementar el conocimiento y contribuir al bienestar y al poder de los pueblos. A la primera se le llama ciencia pura y a la segunda ciencia aplicada. Es claro que esta subdivisión no está delimitada rígidamente: la ciencia es pura solamente en el sentido de que no se preocupa de las aplicaciones, pero sus resultados pueden eventualmente contribuir a la solución de un problema de naturaleza real.

La ciencia es a la vez pensamiento y acción. Se crea ciencia mediante la investigación y su producto final es el conocimiento. La investigación, por su parte, es ante todo un proceso creativo lleno de dificultades e imprevistos. Se construye el conocimiento a medida que avanza el proceso, pero este no es automático: no hay un conjunto de reglas único para tratar un problema científico; es decir, la capacidad

creativa es una función de la imaginación y del talento de una persona, y no puede reducirse a un proceso mecánico.

Se puede, sin embargo, diferenciar dos caminos en la práctica científica: la postulación, que implica la deducción lógica, y la experimentación, que incluye el método científico. El método de la ciencia o método científico, es un proceso que da indicaciones para avanzar en la investigación y suministra medios para evitar algunos errores, pero no reemplaza la inventiva ni la originalidad.

Bunge (1983) describe ocho pasos en la aplicación del método científico, ellos son:

1. Enunciar preguntas bien formuladas y verosímilmente fecundas.
2. Arbitrar conjeturas fundadas y contrastables con la experiencia para contestar a las preguntas formuladas.
3. Derivar consecuencias lógicas de las conjeturas.
4. Arbitrar técnicas para someter las conjeturas a contrastación.
5. Someter, a su vez, a contrastación esas técnicas para comprobar su relevancia.
6. Llevar a cabo la contrastación e interpretar los resultados.
7. Estimar la pretensión de verdad de las conjeturas y la fidelidad de las técnicas.
8. Determinar los dominios en los cuales son válidas las conjeturas y las técnicas, y formular problemas nuevos originados por la investigación.

En un problema o situación real, estos pasos están tan íntimamente relacionados que resulta difícil diferenciarlos completamente. El proceso no es uniforme ni infalible, estando en un paso intermedio es posible que se deba regresar a uno de los pasos iniciales; aun habiendo recorrido todo el proceso, no se garantiza el éxito total. Por otra parte, el desconocimiento o la omisión del proceso científico no invalidan un posible hallazgo o invención, pero sí aumentan terriblemente la probabilidad de fracaso en cualquier intento de contribuir al conocimiento.

1.2 La estadística y el método científico

La experimentación dentro del proceso científico debe ir asistida por técnicas estadísticas: los análisis descriptivo y exploratorio de datos ayudan en el reconocimiento de los problemas y en el planteamiento de las hipótesis, pasos 1 y 2 anteriores; el diseño estadístico de experimentos y el muestreo aportan las bases para la planeación, recolección de datos y validación de la información, pasos 4 y 5; y la inferencia estadística, por medio de la contrastación de hipótesis, estimaciones y conclusiones, pasos 6, 7 y 8, es la única que ofrece métodos basados en la probabilidad para obtener inferencias inductivas válidas.

La estadística sirve de puntal al investigador en el diseño y análisis de sus experimentos. Bien aplicada, conduce a la realización de diseños más eficientes con el consiguiente ahorro de tiempo y recursos, a la vez que se gana en información. Al contrario, experimentos mal diseñados conducen frecuentemente a errores que pueden arruinar todo un proyecto de investigación.

La estadística ayuda a responder preguntas, no a formularlas. Para dar los tres

primeros pasos en la práctica científica, el investigador no requiere la estadística: esta le dice poco acerca de cómo descubrir hipótesis o derivar consecuencias de ellas. Sin embargo, el conocimiento de sus principios y métodos le favorece en el momento de escoger aquellas conjeturas contrastables, que deben traducirse al lenguaje matemático para su verificación.

Por otra parte, mediante la estadística se corrobora o se refuta una hipótesis pero no se demuestra; es decir, esta no puede establecerse con certeza. Aquí se está poniendo en relieve que los métodos estadísticos no son de naturaleza deductiva sino inductiva.

Aún permanece el concepto, no científico, sobre la posibilidad de falsear ciertas conclusiones utilizando estadística. Disraeli, político y escritor inglés del siglo XIX, decía que había tres clases de mentiras: mentirillas, mentiras execrables y estadísticas. Según Di Trocchio (1998), entre las habilidades de algunos científicos para engañar a sus lectores está el conocimiento profundo de los trucos estadísticos, que otorgan la posibilidad de sostener con rigor matemático toda idea surgida de la fantasía. Como ocurre con otras ciencias, la estadística puede utilizarse con fines no siempre plausibles, pero esto no le quita su carácter científico.

1.3 Clases de experimentos

Los experimentos se han clasificado en absolutos y comparativos. En los primeros, el objetivo es la determinación de propiedades absolutas de un conjunto de objetos; por ejemplo, la determinación del número de especies de escarabajos en una región, el porcentaje de individuos que sufren una enfermedad, el peso de un planeta, etc.

En los experimentos comparativos, el objetivo es establecer comparaciones entre objetos que reciben tratamientos diferentes. En agricultura, por ejemplo, es usual comparar semillas, fertilizantes, insecticidas, riegos y suelos; así mismo, la potencia de una droga puede establecerse comparándola con la de otras drogas, incluyendo controles. De igual forma, las diferencias entre individuos, procesos, máquinas, materiales y otros más, se establecen generalmente por comparación.

Los estudios comparativos pueden ser experimentales u observacionales. En los experimentales, el investigador controla, es decir, mantiene constante o varía deliberadamente, aquellos factores que parecen tener mayor influencia en los resultados del fenómeno en estudio. El control de las condiciones experimentales, le permite al científico establecer relaciones causa-efecto entre las variables de respuesta y los factores controlados. A estos también se les llama experimentos diseñados, y son universalmente reconocidos como los métodos explicativos más potentes en la ciencia. Sin embargo, no siempre son aplicables. En medicina por ejemplo, razones éticas y legales podrían impedir su utilización, mientras que en ecología sería necesario esperar cincuenta o más años para comparar experimentalmente el crecimiento de los bosques.

En consecuencia, la estrategia investigativa debe reducirse con frecuencia al enfoque observacional. En este, el investigador no tiene control sobre los factores que causan los cambios en los resultados del fenómeno, y debe limitarse a observar

su forma de manifestarse para establecer relaciones asociativas entre factores y respuestas.

Ambos enfoques investigativos tienen como objetivo dilucidar las relaciones causa-efecto entre las variables en estudio. La metodología estadística aplicada a los experimentos diseñados puede aplicarse también a los experimentos observacionales, aunque las conclusiones son, por lo general, menos convincentes y más débiles para los últimos.

Los estudios observacionales pueden ser descriptivos, exploratorios y analíticos. Los primeros describen el comportamiento de un fenómeno en el tiempo y buscan relaciones o asociaciones estadísticas entre sus variables. Pueden ser transversales si se refieren a un momento en el tiempo, o longitudinales si cubren un periodo establecido para la observación del comportamiento del fenómeno.

Los estudios exploratorios involucran inicialmente un conjunto grande de hipótesis de investigación para, finalmente, desechar aquellas que están en contradicción con los hechos observados y sugerir hipótesis específicas que posteriormente justifiquen una investigación detallada. Por otra parte, los estudios analíticos tienen como finalidad contrastar estas hipótesis particulares.

Ahora bien, los paradigmas observacionales que proceden en la misma dirección que un experimento diseñado, esto es, parten de causas postuladas hacia efectos reconocibles, se llaman experimentos prospectivos o de cohorte. Aquellos que siguen el proceso contrario, los que especifican efectos y buscan las causas posibles, son experimentos retrospectivos, llamados también casos y controles. Tanto los unos como los otros son de aplicación frecuente en epidemología, ecología y en las ciencias sociales, Schlesselman (1982), Grimes y Schulz (2002).

Cuando el científico toma la iniciativa de estudiar un fenómeno aleatorio, ya tiene al frente un problema estadístico; planteado el problema, es necesario recoger información para hacer inferencias. El diseño y análisis de experimentos caen en la esfera de la recolección e interpretación de datos, pero el muestreo también está orientado en la misma dirección. Según Das y Giri (1986), aunque las teorías de encuestas por muestreo y diseño de experimentos conciernen con las técnicas de recolección de datos, cumplen propósitos diferentes. La teoría de encuestas por muestreo deriva métodos para seleccionar muestras de poblaciones con existencia real y casi siempre finitas, de modo que la muestra obtenida representa adecuadamente e interpreta exactamente a la población. En el caso de los experimentos, lo que existe es un problema para resolver y es necesario fabricar o crear la muestra mediante la experimentación apropiada, de modo que la respuesta al problema pueda inferirse de estos datos.

Los métodos de las encuestas por muestreo han sido incluidos por varios escritores como una rama del diseño de experimentos absolutos, en cuanto permite la determinación de características particulares para una población específica (Hinkelmann y Kemptorne, 1994). No obstante, en este texto no se incluirá la teoría del muestreo que por su amplitud puede ser objeto de otro libro, por ejemplo, Díaz (1981), Cochran (1995) y Ospina (2001).

Ejercicios

1.1 Clasificar los experimentos científicos. Sugerencia: hacer un cuadro sinóptico.

1.2 Ubicar la investigación biológica dentro de la ciencia según su objeto, su interés y el camino que sigue en la adquisición del conocimiento.

1.3 Dar ejemplos de experimentos absolutos, estudios experimentales y estudios observacionales.

1.4 Discutir las siguientes afirmaciones:

a. Los experimentos científicos buscan descubrir relaciones causa-efecto entre dos o más variables.
b. Sin la estadística no es posible establecer relaciones causa-efecto entre las variables en estudio.
c. La estadística tiene el peligro de resultar mentirosa si no corrobora las hipótesis de investigación.
d. Con la estadística se puede probar todo y por tanto no se prueba nada.
e. La estadística puede llevar a la certeza en el conocimiento porque se fundamenta en la teoría de la probabilidad.

1.5 Leer el artículo *Introducción al método científico* (Escobar y Urán, 1980).

1.6 Leer el libro escrito por Manotas y otros (1996). Este texto presenta una visión de la metodología de la investigación en las ciencias de la salud.

2

Principios del diseño de experimentos

2.1 Vocabulario del diseño experimental

Muchos de los métodos estadísticos del diseño experimental fueron innovados por R. A. Fisher y F. Yates durante su trabajo en la Estación Experimental Agrícola de Rothamsted, Inglaterra, en los primeros años del siglo XX. De ahí que el vocabulario inicial de esta disciplina se haya derivado de la biología y la agricultura.

Hoy día, el diseño de experimentos se emplea en casi todas las áreas del conocimiento aplicado, y palabras como tratamiento, nivel, bloque y parcela, han perdido su connotación agrícola: su significado original se ha generalizado al aplicarse a la práctica científica de otros campos del saber. Algunos de los términos que se usarán frecuentemente en este libro son los siguientes.

Variables dependientes e independientes. El concepto de independencia en experimentación viene de la idea matemática de relación funcional, donde el cambio de los valores de una variable —dependiente— es el resultado directo de la manipulación de los valores de otra u otras variables —independientes—. Por ejemplo, el peso de un individuo —variable dependiente— varía según su alimentación, edad y sexo —variables independientes.

Variable de respuesta. Es la característica observada o medida en cada unidad experimental; también se llama variable dependiente. En el ejemplo anterior, la variable peso (medida en kg) es la variable de respuesta.

Factor. Es una variable independiente que se evalúa en la investigación. Los factores pueden ser experimentales u observacionales, se llaman factores de estudio y se diferencian de otras variables independientes que no son de interés primordial en el estudio, pero que pueden afectar a la variable de respuesta; estas últimas son llamadas factores de control. Si se investiga el efecto de una dieta en el peso de un individuo, el factor experimental es la dieta, mientras que factores control podrían ser el sexo, el clima, la raza, entre otros.

Niveles de los factores. Un nivel es cada uno de los atributos o estados en que se descompone un factor. Los niveles del factor dieta podrían ser varias clases de dietas; por ejemplo, harina, grasa animal y grasa vegetal. El factor sexo tiene dos

niveles: macho y hembra, y es un factor observacional. La temperatura puede ser un factor de estudio y tomarse a tres niveles, 40, 60 y 80 °C. Una enfermedad puede darse en los estados o niveles inicial, intermedio o avanzado.

Tratamiento. Es un nivel de un factor cuando el estudio es unifactorial. En estudios multifactoriales, un tratamiento es una combinación de niveles tomando uno de cada factor. En el estudio de las dietas, un tratamiento es, por ejemplo, una dieta rica en grasa animal.

Supóngase que se desea investigar el rendimiento de una variedad de avena y se tienen como factores el clima, frío o templado; el riego, sin riego, riego esporádico o riego permanente, y el método de siembra, manual o mecánico. Un tratamiento para este experimento consiste en sembrar manualmente en tierra fría y sin riego. El estudio tendría doce tratamientos en total.

Unidad experimental. Es la parte más pequeña del material experimental a la que se le puede aplicar un tratamiento. Es, por tanto, el objeto sobre el que se realiza una medición u observación.

En agronomía, una unidad experimental puede ser una matera, una parcela, una finca o una región geográfica. En medicina, puede ser una persona, un animal, un consultorio, un hospital, un grupo de diez camas con enfermos, etc. Si se investiga la hipertensión, el grupo sanguíneo, el sida o similares, la unidad es el sujeto. En otros casos, la unidad experimental puede ser el hígado, un miembro o una parte de la piel del individuo.

La unidad experimental también puede ser un sitio escogido en un río, una profundidad dada en un lago, un cuadrante de diez metros en un bosque, un árbol, una rama o una flor de un árbol, etc. Según el objetivo de investigación, una máquina, un lote de materia prima, un día de la semana, un grupo de estudiantes, un hotel, un lugar de la ciudad, también pueden ser unidades experimentales.

En el diseño de un experimento va incluida la labor de definir el tamaño, la forma, el reconocimiento, la ubicación y la diferenciación de las unidades experimentales. Una unidad experimental, en adelante *ue*, puede subdividirse en unidades de muestreo. Si una parcela tiene varios cientos de plantas, la unidad de muestreo podría ser un conjunto de diez plantas seleccionadas al azar. Si la unidad mayor es un estanque con peces, la unidad de muestreo puede ser un pez o un grupo de peces de ese estanque.

Bloque. Es un grupo de unidades experimentales homogéneas. La agrupación de unidades homogéneas en bloques permite reducir el error experimental. Este es uno de los diseños básicos que se verán posteriormente.

Dato. Es el resultado de las observaciones o mediciones hechas a cada unidad experimental y, un conjunto de ellos, conforman la base para hacer los cálculos estadísticos.

Población. Desde el punto de vista del investigador, es el conjunto de todas las unidades experimentales que podrían recibir uno de los tratamientos. El objetivo de la investigación debe establecerse en términos de esta población. Para el estadístico, la población es el conjunto de datos que podrían obtenerse de todas las unidades experimentales.

Para evitar cualquier confusión, algunos llaman universo al conjunto de unidades experimentales y población al conjunto de datos que originarían las unidades experimentales si todas fuesen medidas u observadas. Por ejemplo, las personas mayores de 50 años que viven en un país dado, son el universo para el estudio del peso, y sus mediciones (en kg) constituyen la población. En adelante, se entenderá el término población en ambos sentidos, como unidades o como mediciones de esas unidades.

Muestra. Es un subconjunto de la población que contiene información parcial sobre ella. En experimentos diseñados, las poblaciones son generalmente conceptuales y no siempre es posible muestrearlas según los principios estadísticos teóricos. Por lo anterior, se requiere establecer muy claramente, tanto la población de referencia a la cual se aplicarán los resultados del experimento, como las técnicas que se emplearán en la recolección de la muestra.

Es importante reconocer y diferenciar la terminología del diseño y análisis de experimentos. Para algunos investigadores resulta difícil distinguir entre niveles de factores y variables de respuesta, o entre poblaciones reales y poblaciones conceptuales, y además crean ambigüedades en la definición misma de las unidades experimentales. En un proyecto de investigación no puede haber confusión en la definición de los términos, puesto que la selección del análisis estadístico depende del número de factores por evaluar, del número de niveles de los factores y del tipo de variable de respuesta. Sobre este tópico hay abundantes ejemplos de lectura en Marks (1982).

2.2 Diseño estadístico de experimentos

Según George Wald, citado por Wardlaw (2000), "La experimentación es el mecanismo para hacer que la Naturaleza hable inteligiblemente". Mediante el experimento se pregunta a la naturaleza, pero es necesario diseñar los experimentos para facilitar la comprensión de la respuesta o mensaje implícitos en los datos obtenidos. De otro lado, el análisis estadístico de los datos de un experimento está supeditado al tipo de diseño utilizado; así, diseño y análisis no pueden ir separados en una investigación.

Los diseños experimentales pueden diferenciarse de muchas maneras; John y Quenouille (1977) y Cochran y Cox (1995) traen algunas clasificaciones, pero una enumeración exhaustiva puede resultar en varios cientos de ellos. El objetivo es describir y analizar algunos de los más comunes, y este será el tema central de los próximos capítulos.

Por diseño estadístico de experimentos se entenderá el proceso de planear todos los pasos que deben darse, y el orden que debe seguirse, en la recolección y posterior análisis de la información requerida para estudiar un problema de investigación. En un sentido más restringido, algunos autores consideran el diseño de experimentos como aquel conjunto de reglas que permiten asignar tratamientos a unidades experimentales, buscando eliminar sesgos y cuantificar el error experimental.

El propósito del diseño experimental es la obtención del máximo de información con el mínimo de costo y el máximo de eficiencia. En muchas ocasiones, estas

condiciones se cumplen al seleccionar diseños simples, sencillos de aplicar y fáciles de analizar. Ahora bien, un diseño es eficiente cuando proporciona más conocimientos y mayor precisión con un número menor de datos.

El diseño estadístico de una investigación exige que las personas responsables del experimento tengan una idea clara, desde un principio, de lo que se va a estudiar, de la manera de recoger la información y, al menos, una idea cualitativa del tipo de análisis posterior que se debe hacer a los datos.

2.3 Fases del diseño de experimentos

El diseño y el análisis estadístico de los experimentos se realizan en varias fases; estas han sido descritas y enfatizadas según el propósito de cada autor, entre ellos Anderson y MacLean (1974), John y Quenouille (1977) y Montgomery (2007). Este último autor sugiere las siete fases siguientes:

1. *Reconocimiento y formulación del problema.* Es un paso obvio a primera vista, pero en la práctica no siempre es fácil reconocer un problema que requiere experimentación y, menos aún, formularlo de manera clara y aceptable. La formulación surge de la discusión, generalmente en un comité, donde los integrantes participan con el conocimiento de las posibles causas que originan el fenómeno.

   Para establecer el objetivo general de una investigación y evaluar su importancia, es usual iniciar con una revisión del tema en la literatura. La mejor guía para formular hipótesis se encuentra en los trabajos publicados por otros investigadores. De todos modos, el investigador debería responder breve y claramente preguntas como: ¿Por qué va a efectuar la investigación? ¿Cómo la va a llevar a cabo? ¿A qué tipo de conclusiones espera llegar y qué validez tendrían? ¿A qué población podrían aplicarse las conclusiones? ¿Logrará los propósitos establecidos con los recursos financieros, humanos y de tiempo disponibles?

2. *Selección de los factores y sus niveles.* El investigador selecciona los factores o variables independientes que van a ser analizados, de acuerdo con el conocimiento del fenómeno y la posibilidad de controlarlos. Deberá especificar los factores cuantitativos y los cualitativos, y escoger sus niveles de modo que puedan combinarse y diferenciarse. Dichos niveles pueden ser fijos, o sea, conocidos previamente, o aleatorios, es decir, elegidos al azar de un conjunto mayor. Además, el número de factores y de niveles influyen en el tamaño del experimento y en los métodos de análisis; a su vez, el tamaño es función de los recursos presupuestados.

   Por otra parte, el tiempo puede considerarse como un factor de análisis en el experimento, o como una variable de bloqueo para aislar su influencia cuando no sea de importancia.

3. *Selección de las variables de respuesta.* La variable dependiente o de respuesta debe ser aquella que proporcione información real acerca del problema en estudio. Es necesario, entonces, indicar cómo va a medirse esa variable y qué precisión tendrán las mediciones. Desde luego que las respuestas a estas preguntas están ligadas al tipo de unidad experimental que va a manejarse, a la sensibilidad del instrumento de medida y a la experiencia de quien realiza las mediciones. A su

vez, la unidad experimental la define el tipo de población, espacio inferencial o rango de validez, al cual se extenderán las inferencias y conclusiones finales.

4. *Selección del diseño experimental.* Especificados los niveles de los factores y las variables de respuesta e identificadas las unidades experimentales, surgen dos preguntas relacionadas con el experimento:

a. ¿Cuál es el tamaño apropiado de la muestra? O sea, ¿cuántas unidades experimentales se requieren en el experimento?

b. ¿Cómo asignar los tratamientos a estas unidades experimentales? Es decir, ¿qué tipo de diseño conduce a resultados óptimos en cuanto a información válida, precisa y económica?

Acerca del tamaño muestral, es necesario hacer varias consideraciones propias de la inferencia estadística; este será un tema que se retomará en relación con los diferentes diseños. La segunda pregunta nos lleva a la selección del diseño propiamente dicho. Este determina la manera de seleccionar las unidades y la aleatorización requerida en los tratamientos, para garantizar la ausencia de sesgos en los resultados.

Tres aspectos fundamentales deben tenerse en cuenta en la selección de un diseño: el número de factores, la agrupación de unidades experimentales para eliminar una o más causas de variación, y el número de repeticiones por bloque para definir los diseños de bloques completos e incompletos. Según el diseño elegido, debe proponerse también un modelo matemático para el experimento, de modo que pueda efectuarse un análisis estadístico apropiado de los datos. El modelo muestra la variable de respuesta como una función de los factores en estudio y de algunas restricciones impuestas sobre el experimento por el método de aleatorización.

Un experimento se diseña buscando eficiencia y optimación; sin embargo, también puede realizarse sin ningún diseño previo; la diferencia está en la cantidad de información que puede extraerse de los datos. Algunos investigadores proceden a experimentar sin concebir ningún diseño, con la vaga esperanza de que el análisis estadístico contenga toda clase de magia para extraer más información de los datos de la que en realidad contienen.

Al contrario, un experimento bien diseñado puede producir ganancias, pues las conclusiones que se esperan pueden resultar tan evidentes que hacen innecesario un análisis laborioso de los datos.

5. *Realización del experimento.* Es el proceso de recolección de datos; incluye el control del plan trazado inicialmente y debe prestarse atención al mecanismo de aleatorización, al manejo de los instrumentos de medida, al reconocimiento de las unidades experimentales y, fundamentalmente, al mantenimiento en la forma más uniforme posible de las condiciones ambientales del experimento. Todo el equipo humano debe ser experto en sus tareas y, aunque el experimento no sea complicado, cada uno debe tener un formato específico para ser llenado con los datos y la información pertinente a su obtención. Cuando hay poca experiencia, se recomienda realizar un *pretest* del plan experimental y de la forma de recoger la información, usando unas cuantas unidades experimentales para este fin.

6. *Análisis de los datos.* Los experimentos generan casi siempre un número

menor de datos que las encuestas, pero su análisis no es simple y debe realizarse siguiendo un procedimiento estadístico acorde con el diseño utilizado. La mayoría de los cálculos estadísticos pueden efectuarse con una calculadora de bolsillo, aunque el trabajo puede volverse tedioso. Afortunadamente, hoy día el microcomputador ofrece al analista mayor exactitud numérica, rapidez y, sobre todo, una visión gráfica indispensable en la exploración de los datos. Sin embargo, el computador no está programado para tomar decisiones sino para construir gráficos y tablas y hacer cálculos numéricos, y las conclusiones que se obtengan de estos son responsabilidad de las personas que procesan la información y no de la máquina utilizada.

A menos que se haya seguido un orden en la experimentación, puede ocurrir que mientras más datos se tengan mayor sea la confusión y el desespero durante el análisis. Los estadísticos no se cansan de recalcar que el análisis de los datos no puede separarse del diseño, o sea, de la forma como estos fueron tomados. Finney (1963) sostiene como principio general que "el éxito de un experimento al responder las preguntas que interesan al experimentador, sin demasiados gastos de tiempo y recursos, depende en gran medida de la correcta escogencia del diseño". En este sentido, podrían evitarse problemas serios en el análisis de datos si el investigador revisa y tiene en cuenta los principios básicos antes de realizar el experimento.

Se ha presentado la situación del investigador que gasta meses recogiendo datos y que, finalmente, recurre al asesor estadístico preguntándole qué puede hacer con ellos y pidiéndole que le entregue el análisis de los datos y la respuesta en el menor tiempo posible. El asesor se enfrenta al hecho de compartir la culpa de un estudio mal planeado o de descartar los métodos estadísticos en el análisis de dicho estudio. En ambos casos, queda claro que hubo derroche de los recursos de la investigación. La anterior era una situación más o menos generalizada, como lo observa Gill (1978) cuando escribe: "en el pasado, pero también recientemente, el análisis estadístico a menudo ha sido interpretado como la manipulación de datos ambiguos usando métodos dudosos para resolver un problema que no ha sido definido".

Por otro lado, no es raro escuchar que mediante el computador es posible analizar todo tipo de datos, incluso datos de dudosa procedencia. Pero también se escucha la respuesta del estadístico: si al computador le entra basura, basura imprime. A manera de moraleja, con referencia al análisis de datos, decía Mark Twain: "los datos estadísticos son como la basura, uno debe tener en la mente lo que piensa hacer con ella antes de recogerla".

7. *Conclusiones y recomendaciones.* Después de analizar los datos se establecen las inferencias estadísticas. Estas deben traducirse al lenguaje del experimento para darles la correspondiente interpretación física y evaluar su significación práctica. Los hallazgos deben ir complementados con recomendaciones para estudios posteriores.

Un buen experimento debe generar más preguntas de las que responde, contribuyendo así al proceso iterativo en la adquisición del conocimiento. Algunas recomendaciones deberán presentarse como un informe escrito dirigido a la

persona que debe tomar las decisiones. La cuestión de seguir dichas recomendaciones es, por supuesto, una decisión administrativa; ahora bien, con base en resultados estadísticos, el administrador puede tomar decisiones inteligentes.

Generalmente, el trabajo del asesor estadístico no termina con el análisis de los datos, él debe, en estrecha asociación con el investigador, prepararse para dar sentido a las conclusiones en el contexto del campo científico del experimento.

En conclusión, las primeras cuatro fases exigen el trabajo mental del experimentador y son fundamentales para el éxito del proyecto. Si el investigador piensa utilizar métodos estadísticos para analizar los resultados del experimento, deberá, en primer lugar, identificar esos métodos y orientar el proyecto hacia sus exigencias, antes de realizar el experimento. Por otro lado, la recolección de datos —fase cinco— no debería iniciarse antes de haber respondido satisfactoriamente algunas preguntas como:

a. ¿Cuál es el objetivo principal del proyecto de investigación?

b. ¿Qué tipo de datos van a obtenerse?

c. ¿Qué diferencias en los resultados promedio de los tratamientos se consideran realmente importantes?

d. ¿Qué proceso de selección será usado en la escogencia de las unidades experimentales?

e. ¿Podrán recibir las unidades experimentales más de un tratamiento?

f. ¿Cómo se asignarán los tratamientos a las unidades experimentales?

g. ¿Cuántas unidades experimentales serán necesarias en el proyecto?

h. ¿Cuál será la técnica estadística apropiada en el análisis de los datos?

2.4 Principios básicos del diseño de experimentos

La mayoría de los conceptos sobre planeación y diseño de experimentos discutidos hasta ahora son más de naturaleza metodológica y filosófica que estadística. A partir de este punto, es necesario considerar algunos requisitos del diseño que tienen su origen en las bases teóricas de la estadística inferencial. Posiblemente, los siguientes conceptos no sean ideas estadísticas elementales, se mencionan aquí porque son la esencia del diseño y con el propósito de ampliarlas y darles mayor claridad en páginas posteriores.

Para que un experimento pueda considerarse verdaderamente científico y catalogarse como un buen experimento, las conclusiones obtenidas de este deben tener validez, precisión y amplio cubrimiento.

Una conclusión de origen experimental es válida si está sustentada por técnicas estadísticas probadas anteriormente, y por datos que satisfagan los requisitos teóricos de esas técnicas. Hay dos exigencias teóricas sobre los datos: la ausencia de errores sistemáticos o sesgos y la independencia probabilística entre ellos. Si se garantizan estas condiciones, pueden obtenerse estimaciones insesgadas, tanto de los efectos promedio de los tratamientos como del error experimental. Como ya se ha dicho, la clave para que un diseño origine buenos datos está en la asignación correcta de los tratamientos a las unidades experimentales, *ue*, y la mejor manera de hacerlo es mediante la aleatorización. La idea de aleatorizar es absolutamente fundamental

en el diseño de experimentos válidos y ha sido una de las mayores contribuciones hechas por R. A. Fisher a la metodología científica.

El concepto de precisión en estadística se usa en dos situaciones que es necesario diferenciar. Se habla de precisión de los datos y de precisión de la estimación de un parámetro o de la prueba de una hipótesis. En el primer caso, la precisión se mide por la varianza de los datos; mientras que en el segundo, la precisión es una función del error experimental y este depende no solo de la variabilidad de la *ue*, sino también del número de repeticiones de los tratamientos. Las repeticiones no influyen en la precisión de las mediciones de la variable dependiente, pero son extremadamente importantes cuando se trata de aumentar la precisión de las estimaciones y de las conclusiones del experimento. La precisión en los datos es el resultado de un buen control en el proceso experimental, mientras que la precisión en las conclusiones requiere, tanto la precisión en los datos como la repetición de los tratamientos.

Como ya se dijo, el error experimental —concepto que se definirá posteriormente— depende de la variabilidad de las condiciones experimentales. Su magnitud será menor mientras más homogéneas sean las *ue* que reciben el mismo tratamiento, pero demasiada homogeneidad podría restringir el alcance de la inferencia. Se han ideado varios diseños experimentales para tratar de resolver el conflicto entre amplio cubrimiento poblacional y una precisión deseable, con el objetivo de reducir el error experimental.

En conclusión, las condiciones de validez, precisión y generalidad enumeradas anteriormente, se satisfacen si el experimento se lleva a cabo teniendo en cuenta los tres principios básicos: aleatorización, repetición y control del error experimental.

2.5 Aleatorización

Supóngase que se realiza un experimento para comparar dos dietas en ratas: la dieta A se da a los machos y la dieta B a las hembras. En este experimento no es posible separar el efecto conjunto de la dieta y el sexo, o sea, el resultado del tratamiento está confundido entre los dos factores. Si se concluye que hay diferencias significativas, estas no pueden atribuirse exclusivamente a las dietas, ya que el factor sexo podría haber influido en dicha diferencia, además, con el diseño utilizado no es posible reconocer cuál de los dos factores fue el causante de la diferencia.

En el caso en que ambos factores tengan influencia, tampoco es posible saber en qué proporción contribuye cada uno a esa diferencia. El efecto de la dieta se confunde con el efecto del sexo —suponiendo, obviamente, influencia del sexo—, lo que sesga e invalida cualquier conclusión del experimento.

Si un tratamiento se asigna en la mañana y otro en la tarde, o si se aplica un procedimiento alternado; es decir, a la primera unidad se le asigna el tratamiento A, a la segunda el B, a la tercera otra vez el A, a la cuarta el B, y así sucesivamente, se estarían introduciendo formas sistemáticas de experimentación corriendo el riesgo de sesgar los resultados y, por supuesto, de invalidar las conclusiones finales. Una manera de evitar los sesgos, tanto en los datos como en las estimaciones que se derivan de ellos, es la aleatorización. La necesidad de aleatorizar en la investigación

cuantitativa es reconocida mundialmente y varios autores como Greenberg (1951), Brownlee (1957), Youden (1972) y Fisher (1974) la sustentan explícitamente, aunque otros han planteado controversia en favor de la asignación sistemática, como Gosset (1937).

En el vocabulario del diseño, aleatorizar es asignar el tratamiento a la unidad experimental que lo va a recibir, utilizando un mecanismo de azar; este puede ser una tabla de números aleatorios, monedas, dados o un listado de números seudoaleatorios producido por una calculadora electrónica o un computador. En este sentido, azar y aleatoriedad son sinónimos.

El concepto de *azar* es diferente según el contexto donde se use. Un resultado de azar puede ser un resultado casual, fortuito, inesperado o desorganizado, pero en el sentido probabilístico, indica un resultado, entre varios posibles, que se dio por causas de la naturaleza y sin que hubiesen mediado influencias subjetivas reconocibles.

Por tanto, el experimentador debe adoptar una posición realista ante la aleatorización total o ideal y la asignación sistemática; además, debe estar convencido de que se requiere algún grado de aleatorización para que los análisis estadísticos tengan validez.

No hay un mecanismo único para aleatorizar un experimento, y no todos los diseños exigen la misma aleatorización. Es decir, la aleatorización no es un procedimiento rígido; en algunos diseños puede ser completa, pero en otros solo puede aplicarse parcialmente o en forma restringida, claro, sin llegar al extremo opuesto, en el cual se tendría una asignación sistemática o subjetiva que no es aceptada por la ciencia —la práctica de la aleatorización será explicada más adelante cuando se vean los diferentes diseños.

La aleatorización cumple dos propósitos en el diseño. Por un lado, evita los sesgos causados por fuentes extrañas de variación no controlables y, por el otro, la tendencia, consciente o inconsciente, del experimentador o del proceso, a favorecer grupos únicos de *ue*. Asegura, entonces, que las *ue* que reciben el tratamiento A no difieran sistemáticamente en sus condiciones experimentales de aquellas que reciben el tratamiento B.

Una función de la aleatorización es cancelar los efectos de la variación no controlada que podrían surgir de varias fuentes, principalmente del material experimental y de las etapas de realización del experimento. Cochran y Cox (1995) asemejan la aleatorización a una póliza de seguros en cuanto aquella protege al investigador contra sesgos que pueden no presentarse en los datos, pero que de presentarse ocasionarían resultados funestos.

En el otro contexto, la aleatorización permite suponer independencia de los errores aleatorios en los datos. En experimentos comparativos, el análisis estadístico incluye diferentes pruebas de hipótesis, y la teoría probabilística en que se apoya supone que estos errores están distribuidos independientemente.

Si se supone que de cada *ue* se obtiene un dato, se espera que este difiera de una unidad a otra, ya que existe variabilidad en la población. Por tanto, la mayoría de los datos, si no todos, serán diferentes del valor verdadero del parámetro que representan. Estas diferencias originan los errores asociados con los datos.

No se garantiza independencia en los errores cuando, por ejemplo, algunos

tratamientos se aplican en secuencia a *ue* adyacentes en el espacio o en el tiempo. Aquí, los errores tienden a correlacionarse y si esta correlación o dependencia es positiva, el caso más común, la varianza del error resulta subestimada, lo cual podría invalidar las pruebas estadísticas ya que aumenta la probabilidad del error de tipo I. La correlación negativa casi no se presenta; sus efectos serían los recíprocos, pero aunque menos dañinos, siguen siendo indeseables.

Lo maravilloso de la aleatorización es que destruye la correlación cualquiera que esta sea y produce errores independientes. Esta conclusión la explica Ostle (1988) en los siguientes términos:

La aleatorización nos asegura que el efecto de la correlación sobre cualquier comparación entre tratamientos será mínima. Aún quedará algo de correlación, pero ninguna cantidad de aleatorización podrá eliminarla totalmente. Es decir, en cualquier experimento la independencia de errores completa y verdadera es solo ideal y nunca puede lograrse. Sin embargo, por todos los conceptos, deberá buscarse tal independencia y la aleatorización es la mejor técnica empleada para lograr el fin deseado.

Una conclusión general es que siempre deberá usarse aleatorización en la experimentación y no confiar en una asignación subjetiva, ya que este método tiene desventajas serias. En primer lugar, puede introducir sesgos no detectables en el análisis posterior de los datos; y en segundo lugar, puede originar correlaciones altas, lo cual hace que no se cumplan las suposiciones teóricas de los métodos estadísticos.

Finney (1963), después de discutir la aleatorización, concluye: "Si un experimentador finalmente insiste en que su experimento no puede o no requiere ser aleatorizado, la responsabilidad última será suya; el estadístico puede analizarlo como si hubiese sido aleatorizado, pero este análisis dependerá en todo o en parte, para su validez, del juicio personal y no de la teoría estadística".

2.6 Muestras aleatorias y muestras representativas

El análisis inferencial se aplicará únicamente a muestras aleatorias, es decir, muestras estadísticas obtenidas de una investigación en la que se aplicó algún tipo de aleatorización. Se lee con frecuencia que las muestras deben ser *representativas*, pero este término no está definido en estadística: aleatoriedad no es equivalente a representatividad en la teoría del muestreo. Si por *muestra representativa* se entiende una imagen pequeña de la población en todo sentido, no se sabe cómo seleccionar tal muestra. Algunas muestras aleatorias podrían resultar *representativas* de la población, pero no hay manera de juzgar dicha representatividad.

Aunque el título de esta sección introduce el término en discusión, la intención fue colocarlo allí para llamar la atención del lector más bien que para seguir usándolo en el resto del libro.

Considerando muestras aleatorias únicamente, es posible que se den ciertos arreglos o secuencias en la asignación de tratamientos que resultan inaceptables por lo inesperados y por su apariencia sistemática. Considérese la asignación de tres tratamientos *a*, *b* y *c*, a nueve *ue* en grupos de tres. Utilizando una tabla de números aleatorios, se obtuvo la siguiente secuencia: *acbabccab*. Esta es una de las 1 680 se-

cuencias posibles. Si en vez de la anterior se hubiese tenido la secuencia *aaabbbccc*, claramente habría dudas para aplicarla al experimento, aunque también sea una secuencia aleatoria. Hay tres formas de resolver esta dificultad de acuerdo con Cox (1992): introducir condiciones en el orden de experimentación, rechazar arreglos extremos y realeatorizar o usar una aleatorización restringida del tipo sugerido por Youden (1972).

2.7 Repetición

Si el mismo tratamiento se aplica a *r ue* en un experimento, se dice que el tratamiento está repetido r veces. Si en un diseño cada uno de los tratamientos ha sido repetido *r* veces, se dice que el diseño tiene *r* repeticiones o réplicas. Según Neter y otros (1996), no todas las repeticiones son réplicas; el concepto de réplica es más amplio y se refiere al experimento completo, mientras que el concepto de repetición se aplica a un tratamiento particular. La mayoría de los autores emplean indistintamente los vocablos repetición y réplica, y en este sentido se aceptará aquí, aunque, en adelante, se usará con mayor frecuencia el primero.

Ciertas mediciones hechas sobre la misma *ue* no constituyen verdaderas repeticiones. Por ejemplo, si la unidad de análisis es el individuo y se quiere medir una característica en su sangre, el caso de tres lecturas de la misma muestra —tres submuestras— de sangre no representa repeticiones del tratamiento aplicado al individuo. En este caso, la repetición por submuestreo ayuda en el control y en la estimación de errores analíticos, pero no contribuye a disminuir el error experimental.

Según el enfoque de Steel y Torrie (1997), la repetición de un diseño tiene cuatro funciones:

1. Permitir una estimación de la varianza del error experimental.
2. Mejorar la precisión de las estimaciones en el experimento.
3. Aumentar el alcance de la inferencia.
4. Ejercer control sobre la varianza del error.

Los científicos son conscientes de que es prácticamente imposible explicar todas las fuentes de variación que afectan las mediciones experimentales. La porción de la variabilidad que no es explicable por factores conocidos se llama error experimental, y la forma de medirlo es la varianza del error por unidad experimental, σ^2. Esta varianza es una función de las diferencias entre mediciones hechas a las *ue* sometidas al mismo tratamiento, y del número de repeticiones de los tratamientos. Si no hay repeticiones, la varianza no es estimable. El problema es que las inferencias están basadas en dicha varianza y si esta no puede estimarse de los datos, no es posible obtener conclusiones estadísticas reales.

Ahora bien, la precisión de un experimento se evalúa por sus resultados inferenciales. En primer lugar, mientras más corto sea el intervalo de confianza, mayor es la precisión de la estimación de efectos de tratamientos. Pero la longitud del intervalo depende directamente del error estándar de esta estimación, calculado como $\sqrt{\sigma^2/r}$; así que a mayor número de repeticiones, mayor precisión. En cuanto a las pruebas de hipótesis, su precisión, medida como sensibilidad o potencia, está

relacionada igualmente con el error estándar. Sobre estos temas se volverá en el análisis posterior de algunos diseños. También se incluirá el problema de la estimación del número de repeticiones requeridas en el diseño para lograr una precisión y una confiabilidad preestablecidas.

Uno de los propósitos del diseño es aumentar la precisión, o por lo menos mantenerla constante en un valor prefijado, sin sacrificar inadecuadamente el rango de validez de las conclusiones. De esta manera, mientras más generales sean las condiciones de investigación, mayor es la confianza en la extrapolación de los resultados experimentales. Para generalizar a una población más amplia se requiere mayor número de repeticiones, lo que implica manejar *ue* mucho más heterogéneas. En este caso, el problema de la precisión se controla, en parte, formando grupos o bloques con unidades similares.

En experimentos diseñados, la población de referencia normalmente no puede muestrearse con las técnicas usuales, y el proceso se reduce a la selección de algunas *ue* que estén disponibles. Este procedimiento no es aceptable a causa de que reduce la población objeto de la inferencia. Por ejemplo, muchos ensayos fisiológicos se han realizado en ratas que han sido compradas o criadas en un mismo lugar, de una progenie específica y altamente endogámica. Es poco probable que el experimentador hubiese deseado aplicar sus resultados a una sola línea de endogamia; lo más probable es que no haya definido la población de referencia, aunque intentara seguir el principio estadístico de minimizar la variabilidad.

Para algunos investigadores, la única manera conocida de disminuir la variabilidad era la selección de unidades homogéneas, pero hoy día el diseño experimental permite ampliar la base de la inferencia mediante la formación de bloques, evitando los efectos adversos de variables como la descendencia, el sexo, la edad, el peso, el tiempo, etc. Cuando las *ue* se agrupan en bloques, de modo que la variación entre unidades dentro de cada bloque sea menor que la variación entre unidades en bloques diferentes, la precisión del experimento aumenta como resultado del control en la varianza del error. En este diseño, los bloques completos se llaman también repeticiones y, si su número aumenta, el error experimental será menor.

2.8 Control de error experimental

La estadística debe su existencia a la necesidad de analizar la variabilidad presente en los resultados de un fenómeno. En particular, en los experimentos bien diseñados esta variabilidad ha sido modelada como:

Dato observado = (valor promedio poblacional)
+ (una cantidad debida al tratamiento usado)
+ (una cantidad debida a la unidad experimental).

La cantidad de variación atribuible a la *ue* proviene de dos fuentes: la variabilidad inherente al material experimental al que se le aplican los tratamientos, y la falta de uniformidad en la conducción física del experimento. Por ejemplo, en experimentos con ratas como *ue*, la constitución genética es un factor que origina

variabilidad inherente al material experimental, excepto cuando haya alta endogamia. Las ratas se colocan en jaulas sujetas a diferencias de calor, luz, altura, ruido y otros factores más, lo cual constituye la falta de uniformidad en la realización del experimento.

La variabilidad en las *ue* se debe a dos formas de error, uno sistemático y otro aleatorio. Como se dijo anteriormente el error sistemático surge de un proceso de experimentación deficiente que incluye carencia de aleatorización, instrumentos de medida mal calibrados, personal sin entrenamiento previo, etc. Una forma de evitar estos sesgos es, por tanto, el refinamiento de la técnica experimental. Si se supone que todas estas fallas se han remediado al máximo, aún queda una parte de la variación que no es controlable, ya sea porque proviene de causas desconocidas o porque, habiendo sido reconocidas, su control podría resultar muy complejo o muy costoso. Así, la porción remanente de la variación en la *ue* se atribuye a causas del azar y constituye el error aleatorio.

En un experimento bien diseñado deben eliminarse los sesgos y minimizarse el resto de variación, de tal manera que el error experimental resultante sea únicamente el mínimo error aleatorio que siempre está presente en cualquier investigación experimental.

El diseño experimental se concibe como el proceso que lleva al control del error, entendiéndose como tal, aquellos pasos que conducen a tomar todas las precauciones necesarias para eliminar el máximo número de causas que producen error y reducir al mínimo la influencia de aquellas que no son eliminables. Es decir, en todo experimento hay error; la idea del diseño es reducir su magnitud, pues resulta imposible eliminarlo del todo.

En la tarea de controlar el error experimental, parte corresponde al experimentador y parte a la estadística. En cuanto al primero, los pasos que se deben seguir incluyen:

1. Seleccionar material experimental homogéneo pero no al extremo de restringir la población objeto de la inferencia.

2. Elegir la forma y el tamaño apropiados de las *ue*. En general, las unidades grandes son menos variables que las pequeñas, pero, a la vez, las unidades pequeñas son más fáciles de manipular y pueden repetirse mayor número de veces, tanto en experimentos de campo como en experimentos de laboratorio.

3. Aumentar la magnitud del experimento. Mientras más tratamientos y más repeticiones se planeen, puede esperarse mayor precisión. No hay que olvidar que el tamaño del experimento lo limitan el tiempo y los fondos disponibles.

4. Usar la *ue* como su propio control siempre que sea posible. Aquí la connotación de la palabra control es la de la *ue* no tratada. Un grupo control es aquel que no recibe ninguno de los tratamientos planeados, pero que es necesario incluirlo en el experimento para separar los efectos de los tratamientos de aquellos efectos causados por las otras condiciones experimentales.

5. Refinar la técnica experimental. Recuérdese que ninguna clase de análisis estadístico puede mejorar datos obtenidos de experimentos mal hechos.

Las siguientes son algunas de las pautas más recomendadas en la experimentación:

a. Tener uniformidad en la aplicación de los tratamientos

b. Ejercer control sobre la influencia de variables exógenas o que no entran en el estudio.

c. Usar técnicas de medición adecuadas, aparatos y procesos con precisión garantizada, y personal con entrenamiento y conocimientos reconocidos.

La estadística ayuda en el control del error en dos formas: mediante la escogencia del diseño y sugiriendo mediciones adicionales de variables relacionadas con la variable de respuesta, llamadas covariables. La utilización de estos métodos se conoce como control local.

Una estimación insesgada de la varianza del error experimental se obtiene aplicando una técnica muy elegante llamada *análisis de varianza*. Si en el análisis entran, además, algunas variables concomitantes o covariables, se habla de *análisis de covarianza*. Su objetivo es, igualmente, aumentar la precisión de las estimaciones de los efectos de los tratamientos. A través de las covariables, por ejemplo el peso inicial, la edad, la procedencia, etc., se puede explicar parte de la variabilidad en la variable de respuesta y aislarla de la variabilidad total, reduciendo de esta forma el error experimental.

Una de las varias formas de clasificación de los diseños experimentales tiene en cuenta el vocabulario descrito en la sección 2.1. Así:

1. Según el número de variables de respuesta los diseños pueden ser:

a. Univariados, es decir, una sola variable de respuesta.

b. Multivariadas, o sea, se analizan varias variables de respuesta simultáneamente.

2. Según el número de factores de estudio pueden ser:

a. Unifactoriales, se analiza un solo factor de estudio.

b. Multifactoriales, se consideran varios factores de estudio simultáneamente en el análisis de la información.

3. Según el número de variables o factores de control, los diseños se clasifican en:

a. Diseños completamente aleatorizados. Estos no incluyen ningún factor de control.

b. Diseños de bloques aleatorizados. En estos diseños interviene una sola variable de control, llamada variable de bloques.

c. Diseños de cuadrado latino. Se reconocen porque manejan simultáneamente dos variables de control.

d. Otros diseños incluyen tres o más variables de control y reciben nombres especiales. Cuadrados greco-latinos, etc.

4. Según el número de repeticiones por tratamiento los hay:

a. Balanceados. Todos los tratamientos con igual número de repeticiones. Cada par de tratamientos se compara con la misma precisión.

b. Desbalanceados. Generalmente se presentan al construir bloques. Resultan bloques incompletos balanceados o parcialmente balanceados, se requiere teoría matemática para la construcción de bloques incompletos, Ragha varao (1971).

5. Según la combinación de los niveles de los factores pueden ser:

a. Cruzados. Cada nivel de un factor se combina con los niveles de los otros factores de estudio.

b. Anidados. Los niveles de un factor se encuentran dentro de los niveles de otro factor. Son diseños con submuestreo.

6. Según la forma de selección de los niveles de cada factor resultan los siguientes modelos:

a. Modelo I o de efectos fijos. Los niveles de cada factor están predeterminados o los fija el investigador.

b. Modelo II o de efectos aleatorios. También se le dice modelo de componentes de varianza. Se caracteriza porque los niveles de los factores son seleccionados al azar entre un grupo grande de niveles o una población.

c. Modelo III o mixto. Incluye factores de efectos fijos y factores de efectos aleatorios.

2.9 Exactitud y precisión

El concepto de precisión no debe confundirse con el de exactitud en las mediciones. La precisión se refiere a la variabilidad de los datos muestrales con respecto a su valor promedio, mientras que la exactitud tiene que ver con la variabilidad respecto a un valor promedio teórico. El ejemplo de Wardlaw (2000) es particularmente ilustrativo: un tirador a un blanco tiene precisión si todos sus disparos dan en puntos cercanos entre sí, y tiene exactitud si sus disparos dan en el blanco o cerca de él.

Se dice que hay precisión cuando todos los resultados son similares entre sí, esto es, son reproducibles; una varianza pequeña indica alta precisión. Hay exactitud cuando los resultados están cercanos al valor verdadero o de referencia; la diferencia con respecto al valor verdadero es el sesgo o error sistemático. De otro lado, datos sesgados son datos inexactos aunque podrían ser precisos.

Lo ideal en un experimento es minimizar el sesgo y aumentar la precisión en los datos para obtener conclusiones más contundentes. La estadística práctica asume que los datos están libres de error sistemático —estos se han eliminado o controlado— y que la variabilidad restante se debe únicamente a errores aleatorios que también han sido minimizados.

Bajo la suposición de ausencia de sesgos, precisión es igual que exactitud, o mejor, incrementar la precisión implica aumentar en exactitud. Si el sesgo es grande, aunque incremente la precisión, el aumento en la exactitud podría ser insignificante. Los sesgos deberán evitarse desde un principio puesto que afectan el error experimental y no son detectables al analizar los datos. Mediante el error experimental puede juzgarse la precisión mas no la exactitud.

En la estimación de la varianza del error experimental, no es posible discriminar o aislar la parte asignable al sesgo de la parte que es el error aleatorio verdadero. Observando el coeficiente de variación, el analista puede hacer comparaciones con resultados de otros investigadores y darse cuenta de si hay grandes fluctuaciones en sus datos. Una comparación desfavorable indica que probablemente hubo demasiadas perturbaciones durante el periodo y ambiente experimentales.

Ejercicios

2.1 Una fábrica está planeando consumir un combustible más barato, pero este libera mayor cantidad de SO_2 al aire que el combustible usado corrientemente. La Oficina de Protección Ambiental no aprueba el cambio de combustible hasta que la empresa compruebe que la cantidad de SO_2 liberada no afecta la respiración de las personas. La situación anterior origina un proyecto de investigación para determinar si la respiración del aire contaminado con SO_2 es perjudicial para las personas con problemas respiratorios o sin ellos.

a. Formular el problema, incluyendo el objetivo general del proyecto de investigación.
b. Especificar la población objeto de estudio, la unidad experimental y un procedimiento factible de obtener la muestra.
c. Elegir posibles factores de estudio y sus niveles. Clasificarlos en cualitativos, cuantitativos, experimentales y observacionales.
d. Elegir la o las variables de respuesta y especificar la forma, unidades de medida e instrumentos para medirlas.
e. Plantear algunas hipótesis verificables con los datos que van a obtenerse.
f. Indicar la aplicabilidad de cada uno de los principios básicos del diseño a este proyecto.

2.2 Formular el problema de investigación que usted siempre ha soñado y que no ha realizado, posiblemente por falta de recursos económicos o de tiempo. Responder preguntas similares a las del proyecto anterior.

2.3 George y otros (1991) presentan en el capítulo 3 de su libro, los pasos que se deben seguir en los diseños de estudios de evaluación ambiental. Se recomienda la lectura de este capítulo, para formarse un concepto sobre la aplicación directa de las fases del diseño en un estudio de tipo biológico.

2.4 Cuáles de los principios básicos de una investigación se requieren para responder correctamente cada una de las preguntas siguientes:

a. ¿Son válidos los resultados de la investigación?
b. ¿Qué precisión tienen los resultados obtenidos? ¿Es suficiente dicha precisión de acuerdo con la literatura científica?
c. ¿Son útiles y generalizables estos resultados?
Justificar las respuestas con explicaciones convincentes.

3

Análisis estadístico de una muestra

3.1 Clasificación de las variables de respuesta

Después de haber establecido los objetivos, el investigador deberá decidir sobre la manera de alcanzarlos. El paso que sigue incluye la selección de las variables de respuesta apropiadas; las contestaciones a las siguientes preguntas ayudan en esta selección:

1. ¿Cuál es la variable de respuesta más apropiada para medir el fenómeno de interés y que permite, a la vez, alcanzar el objetivo propuesto?
2. ¿Qué variable de respuesta puede medirse en la práctica?

En algunos estudios la variable por medir es casi obvia: el peso o volumen de la producción al ensayar fertilizantes en cultivos, el número de especies de plantas o aves en un bosque, la concentración de metales pesados en el agua para consumo humano, etc. Si el objetivo es comparar mortalidad entre las poblaciones, la edad posiblemente no es relevante para este propósito.

Supóngase que un objetivo es determinar si el SO_2 en el aire influye en la respiración de las personas. Aquí se tienen varias formas de medir la variable de respuesta; por ejemplo, mediante la capacidad vital, la capacidad residual funcional o la rata de flujo expiratorio maximal promedio. La elección del método estaría supeditada a la disponibilidad de los aparatos de medida y al manejo práctico del fenómeno.

La variable de respuesta o característica en estudio se analiza como una variable de tipo matemático, llamada *variable aleatoria* o *variable estadística.* Estas variables se clasifican en cuantitativas y cualitativas, como se muestra en la figura 3.1.

Los valores numéricos de las variables tienen relación con las escalas usadas para hacer la observación. Hay cuatro tipos de escalas de medida: nominal, ordinal, de intervalos y de razones. La nominal es la más débil de todas, diferencia objetos o eventos con base en un nombre. La escala de razones es la más precisa y sofisticada, representa el nivel de medición más elevado conocido.

Cada escala aporta una cantidad de información sobre la variable en estudio. Mientras más alto sea el nivel de cuantificación, mejor será la discriminación entre los objetos medidos y, por tanto, habrá mayor objetividad en la información. De ahí que

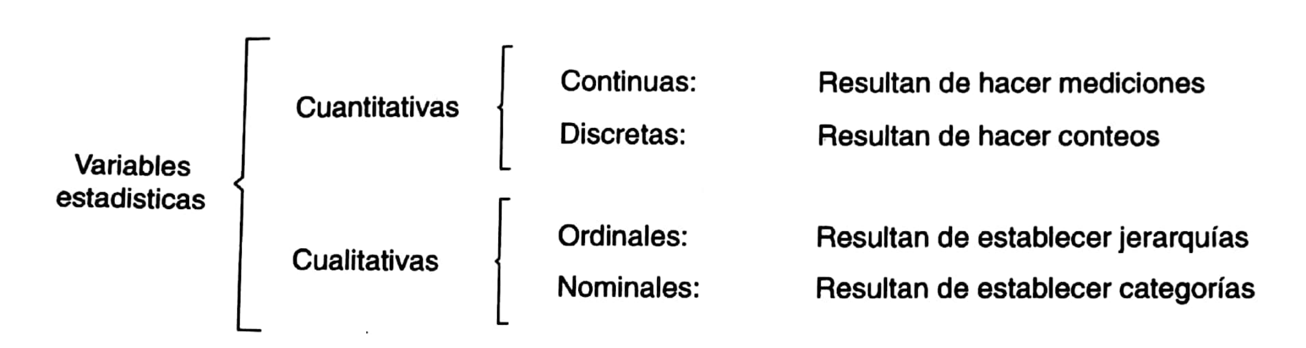


Figura 3.1 Clasificación de las variables estadísticas

todas las disciplinas científicas busquen la cuantificación máxima de sus variables. La mayoría de los procedimientos estadísticos desarrollados para manejar información empírica, dependen del tipo de variable y de la escala usada en su medición. Los métodos de mayor sensibilidad son precisamente aquellos que manejan variables medidas en escala de razones.

3.2 Análisis de una variable cuantitativa

El análisis se realizará mediante un ejemplo y utilizando una variable medible.

Ejemplo 3.1 Durante el primer semestre de 1982, se les pidió a los alumnos de un curso de estadística que tomaran una muestra de las estaturas de los estudiantes de la Universidad de Antioquia, limitándose a la ciudad universitaria. La información fue recogida mediante entrevista personal en diferentes sitios de la Universidad; la estatura se registró en metros y centímetros. De una muestra original de 400 datos, se seleccionó aleatoriamente una submuestra de 40, con la única finalidad de mostrar el análisis estadístico de una variable cuantitativa continua como la estatura.

Se presentará paso a paso el análisis descriptivo, exploratorio, inferencial y de ajuste a la distribución Normal, haciendo énfasis en algunos conceptos que posiblemente no son estudiados en un primer curso de estadística.

1. Datos originales. La tabla 3.1 corresponde a las estaturas en metros de 40 estudiantes de la Universidad de Antioquia seleccionados al azar en la ciudad universitaria; en adelante, estaturas de estudiantes.

2. Medidas descriptivas de la muestra. La estatura es una característica que varía de una persona a otra y puede analizarse como una variable aleatoria continua. Esta se representará con la letra mayúscula Y; en tanto que los valores de la muestra serán representados por la letra minúscula y_j, con $j = 1, 2, \dots n$, donde n es el tamaño de la muestra. Aquí se calculan las siguientes medidas:

a. Media aritmética:

$$\bar{y} = (1/n)\sum_j y_j = 66.21/40 = 1.66 \text{ m.}$$

Este valor es la estatura promedio de los estudiantes en la muestra.

b. Rango:

R = valor máximo – valor mínimo = 1.93 – 1.48 = 0.45 m.

Tabla 3.1 Estaturas de estudiantes de la Universidad de Antioquia

1.50	1.75	1.61	1.75	1.58	1.55	1.59	1.75	1.86	1.93
1.60	1.70	1.58	1.50	1.61	1.58	1.73	1.58	1.52	1.48
1.65	1.53	1.57	1.64	1.72	1.80	1.78	1.74	1.55	1.48
1.58	1.70	1.75	1.62	1.72	1.83	1.80	1.65	1.65	1.70

c. Varianza

$$s^2 = \left[n\sum_j y_j^2 - \left(\sum_j y_j \right)^2 \right] \Big/ n(n-1) = 0.0124 \text{ m}^2.$$

Aquí m^2 no indica área sino unidades al cuadrado, que son las unidades de la varianza.

d. Desviación estándar

$$DE = \sqrt{s^2} = +s = \sqrt{0.0124} = 0.11 \text{ m}.$$

e. Coeficiente de variación

$$CV = (s/\bar{y})100\% = 6.7\%$$

f. Error estándar de la media

$$EE = s\big/\sqrt{n} \doteq 0.02 \text{ m}.$$

Generalmente, la información se reporta en una de las siguientes formas:

como $\bar{y} \pm DE(n)$, o sea, 1.66 ± 0.11 (n = 40), o

como $\bar{y} \pm EE(n)$, o sea, 1.66 ± 0.02 (n = 40).

El intervalo alrededor de la media permite juzgar el grado de precisión de los datos. No obstante, no puede darse un juicio absoluto; este depende de un patrón de referencia ya establecido o es el resultado de la comparación con otros intervalos de estudios similares.

En la primera forma, se puede decir que aproximadamente el 67% de los estudiantes de la muestra tienen una estatura comprendida en el intervalo (1.55, 1.77) m. Multiplicando el *EE* por dos en el segundo caso, obtendremos el intervalo (1.62, 1.70) m, que se puede interpretar como uno de los intervalos del 95% de confianza que contiene la estatura promedio poblacional, suponiendo, en ambas situaciones, que las estaturas tienen distribución Normal en la población. Si se reporta la *DE* se está enfatizando sobre la distribución de los datos en la muestra. En cambio con el *EE* se especifica un intervalo donde puede situarse la media poblacional.

g. Simetría y curtosis. Son medidas que describen la forma geométrica de la distribución de los datos. La forma de dicha distribución da información útil para compararla con el modelo probabilístico Normal, en adelante modelo Normal, empleado en la Inferencia Estadística. Dos de estas medidas son el *coeficiente de asimetría* que mide la cantidad de sesgo a la derecha o a la izquierda de la media de la distribución; es decir, la cola más larga de la curva, y el *coeficiente de curtosis* que permite apreciar el alejamiento en cuanto a la flacura o gordura de la curva de frecuencias con respecto a la curva Normal. Estos coeficientes son de interés cuando el número de datos es grande (n > 100) pero, en este caso, su cálculo es dispendioso y es aconsejable el uso

del computador. Actualmente, se prefiere hacer un análisis exploratorio de los datos en vez de especular con ambos coeficientes.

3. Gráficos del análisis exploratorio de datos. La estadística descriptiva proporciona algunas indicaciones para tabular —tablas de frecuencias— y graficar —histogramas— un conjunto de datos. El objetivo de la graficación es mostrar la forma de la distribución de los datos, pero su éxito depende de la escogencia de los intervalos de clase y, generalmente, en este proceso se pierde información. Tukey (1977), desarrolló algunos métodos gráficos que corrigen las fallas anteriores; estos son más simples de construir y ayudan al analista en el reconocimiento de ciertos patrones de distribución necesarios en la inferencia. Estas técnicas pertenecen al análisis exploratorio de datos, uno de los enfoques más recientes en el análisis de la información estadística. Dos de estos elementos son el gráfico de *ramas* y el gráfico de *cajas esquemáticas.*

3.3 Gráfico de ramas

El gráfico de ramas o traducido literalmente, gráfico de tallo y hojas, es un esquema parcialmente tabular y parcialmente gráfico, que reemplaza en parte el histograma y en parte la tabla de frecuencias (véase figura 3.2, a y b). Cada dato se descompone en dígitos a la izquierda, los cuales forman los tallos, y dígitos a la derecha, estos forman las hojas. Los *tallos* se colocan en una columna y se traza al frente una línea vertical para separarlos de las *hojas* que van horizontalmente al frente de sus respectivos tallos.

Ejemplo 3.2 Tómese la muestra de las estaturas para ilustrar el proceso de construcción de este gráfico y darle sentido al vocabulario utilizado. Los datos varían entre 1.48 y 1.93 m. Se pueden tomar los dos primeros dígitos como los tallos y el último como la hoja; por ejemplo, el dato 1.48 se representará como 1.4|8 y el dato 1.75 como 1.7|5. Los números 1.4, 1.5, ..., 1.9 son los tallos y la última cifra de cada dato representa la hoja. Cada tallo define una clase o intervalo de la distribución; así, 1.5 define al intervalo 1.50 a 1.59.

Las hojas de cada rama forman una figura similar a un histograma girado 90° a la derecha. En el gráfico aparecen todos los datos en forma ordenada para facilitar la localización de la mediana, los cuartiles, etc. Cuando se tienen datos extremos o *outliers* (véase sección 3.8), estos aparecerán en la parte superior o inferior del gráfico para llamar la atención sobre ellos, pues deberán investigarse con mayor cuidado.

A veces el número de tallos no permite mostrar una distribución adecuada de los datos y, como en el ejemplo anterior, es preferible aumentar este número. En vez de tomar tallos de 0.1 en 0.1 se toman de 0.05 en 0.05, haciendo la siguiente subdivisión: un * comprende hojas del 0 al 4 y un $_0$ comprende hojas del 5 al 9, figura 3.2 (b). La columna de la izquierda encabezada con el símbolo #, muestra la frecuencia acumulada de arriba hacia abajo y de abajo hacia arriba.

El gráfico de ramas permite explorar la distribución de un conjunto de datos y detectar algunos comportamientos, como:

1. Forma y grado de simetría.
2. Rango y grado de dispersión.
3. Existencia de valores *outliers.*
4. Intervalos de concentración de algunos datos.
5. Espacios intermedios vacíos o con pocos datos.
6. Localización de la mediana, los cuartiles, el máximo y el mínimo. En la figura 3.2 (b) se observa una leve asimetría positiva, o sea, una cola mayor hacia valores altos, no hay demasiada dispersión, no hay intervalos vacíos, no hay valores excesivamente extremos y la distribución no parece alejarse demasiado del modelo Normal.

1.4	88
1.5	0023557888889
1.6	01124555
1.7	000223455558
1.8	0036
1.9	3

a

#		
2	1.4o	88
6	1.5*	0023
15	1.5o	557888889
20	1.6*	01124
20	1.6o	555
17	1.7*	0002234
10	1.7o	55558
5	1.8*	003
2	1.8o	6
1	1.9*	3

b

Figura 3.2 (a) Gráfico de ramas para el ejemplo 3.2; (b) Gráfico de ramas para el ejemplo 3.1.

3.4 Cajas esquemáticas

Una caja esquemática es una representación gráfica rectangular de la distribución de los datos muestrales; es complementaria del gráfico de ramas. El gráfico de ramas describe, de manera simple y elegante, la forma de la parte central de la distribución de frecuencias. En contraste, la caja esquemática enfoca la atención especialmente sobre las colas de la distribución, pero sin descuidar indicaciones de localización, como la mediana, y de escala, como el rango intercuartil.

Ejemplo 3.3 Con los datos del ejemplo 3.1, se verán los pasos y cálculos requeridos en la construcción de una caja esquemática (véanse figuras 3.3 y 3.5):

1. Se ordenan los datos y se localiza la mediana, *Me,* y los cuartiles inferior, Q_1, y superior, Q_3.

Teniendo los datos ordenados en el gráfico de ramas, se procede así: a cada valor extremo se le asigna una *profundidad* igual a 1; es decir, su localización como primer dato en orden ascendente o descendente.

Profundidad de la $Me = (n + 1) / 2 = (40 + 1) / 2 = 20.5$; entre 20 y 21.

Las posiciones 20 y 21 están ocupadas por los datos 1.64 y 1.65, así que la *Me* es igual a 1.645 m.

Profundidad del cuartil = 1/2 [profundidad truncada de $Me + 1$] = $1/2(20 + 1) = 10.5$.

Las posiciones 10 y 11, en orden creciente, corresponden a los datos 1.58 y 1.58, luego $Q_1 = 1.580$.

Las posiciones 10 y 11, en orden decreciente, corresponden a los datos 1.75 y 1.74, de modo que $Q_3 = 1.745$.

Los valores de Q_1 y Q_3 son aproximados; Tukey los llama *cuartos* en vez de cuartiles.

2. Se calcula el rango intercuartil $Q = Q_3 - Q_1 = 1.745 - 1.580 = 0.165$.

3. Se calcula un factor de escala llamado *paso*, cuyo valor es $P = 1.5\,Q = 1.5 \times 0.165 = 0.2475$.

4. Se calculan dos valores que se llamarán *barreras o límites* inferior y superior, así;
barrera inferior: $B_i = Q_1 - P = 1.5800 - 0.2475 = 1.3325$,
barrera superior: $B_s = Q_1 + P = 1.7450 + 0.2475 = 1.9925$.

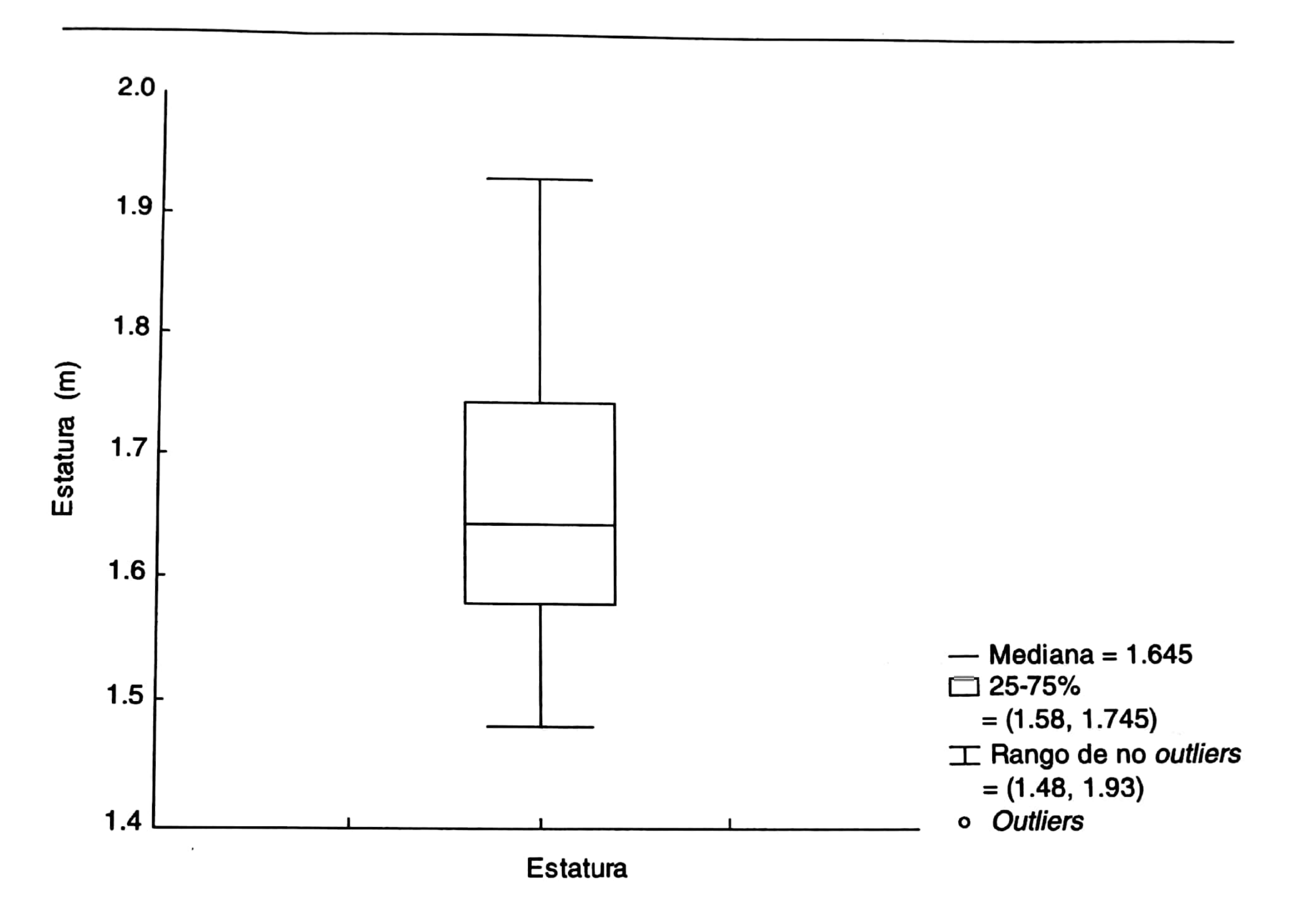


Figura 3.3 Caja esquemática para el ejemplo 3.3

5. Se localizan los dos valores *adyacentes* inferior y superior. Un valor adyacente es el valor, perteneciente al intervalo, más cercano a su respectiva barrera.

Valor adyacente inferior: A_i = 1.48. No siempre es el valor mínimo.

Valor adyacente superior: A_s = 1.93. No siempre es el valor máximo.

6. Los datos ubicados por fuera del intervalo con las barreras como límites, se catalogan como datos *outliers.* Si se amplían las barreras duplicando el paso y aparecen datos por fuera de este intervalo, estos serán definitivamente extraños. Posiblemente serán excluidos de la muestra si se desea un ajuste de los datos al modelo Normal.

7. Se traza una escala vertical que cubra el rango de los datos, y a la derecha se construye un rectángulo de largo igual al rango intercuartil sin importar su anchura. El rectángulo se divide con una línea a la altura de la mediana y desde el centro de los lados inferior y superior se trazan líneas verticales hasta los valores adyacentes respectivos. Los valores *outliers* se representan por puntos individuales más allá del punto extremo de la línea, diferenciando los datos *outliers* de los datos extraños, por ejemplo usando puntos y cruces.

Ejemplo 3.4 Los siguientes datos representan longitudes en centímetros del rizoma de plantas (Melastomataceae) estudiadas por A. Uribe y L. Albert (1991, comunicación personal) en un bosque pluvial premontano del municipio de Guatapé, Departamento de Antioquia (Colombia).

16	8	14	30	7	8	38	2	8	7	16	20	8	10	10	6	5	5	8	8
4	2	5	7	8	3	10	7	4	6	8	4	12	14	9	4	6	3	2	14
8	6	16	8	6	9														

El gráfico de ramas para estos datos (véase figura 3.4) presenta un tallo con demasiadas hojas, lo que sugiere otra subdivisión de los tallos para que los datos no queden concentrados en el segundo intervalo. Obsérvese que el gráfico muestra asimetría positiva alta: los intervalos mayores están vacíos o tienen muy pocos datos, indicando que son valores que deben analizarse individualmente. Aun con una agrupación diferente, los datos grandes sobresalen en el grupo, dicho comportamiento es lo que se quiere resaltar con este ejemplo.

Para construir una caja esquemática se hacen los siguientes cálculos:

#		
9	0*	222334444
33	0_0	555566667777888888888899
13	1*	0002444
6	1_0	666
3	2*	0
2	2_0	
2	3*	0
1	3_0	8

30 y 38 son valores muy altos

Figura 3.4 Gráfico de ramas para el ejemplo 3.4

$Me = 8, Q_1 = 5, Q_3 = 10, Q = 5, P = 7.5,$
$B_i = 5 - 7.5 = -2.5,$

$B_s = 10 + 7.5 = 17.5, A_i = 2, A_s = 16.$

La figura 3.5 muestra cierta simetría en los datos si se excluyen los dos mayores, que parecen definitivamente extraños.

Las cajas esquemáticas muestran diferentes características de la distribución de los datos. En ellas es posible reconocer fácilmente:

1. Un valor central, o sea, la línea de la mediana, y la posición de los cuartiles.

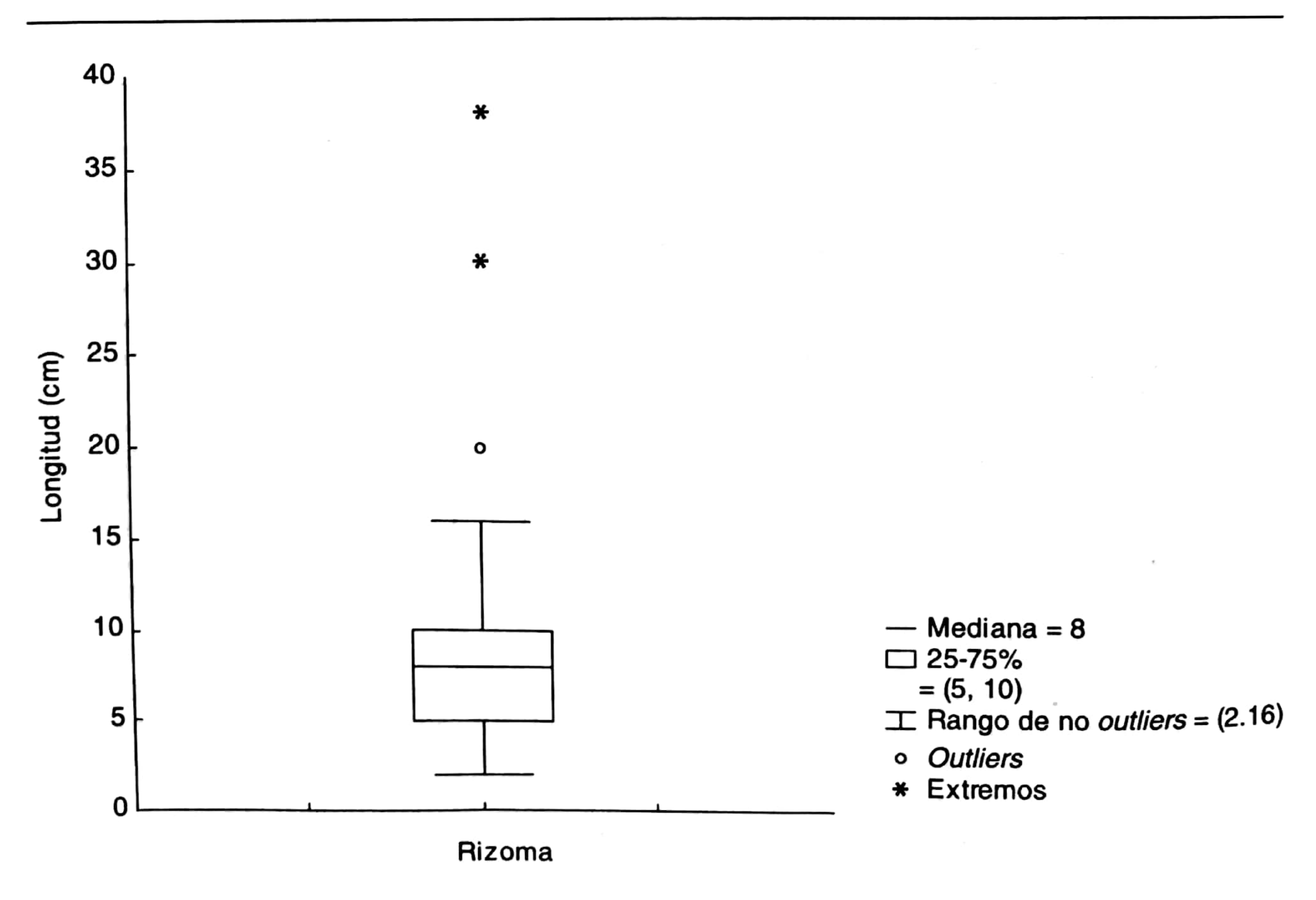


Figura 3.5 Caja esquemática para el ejemplo 3.4

2. La dispersión del 50% de los datos centrales o rango intercuartil, representado por la longitud vertical del rectángulo.

3. La asimetría de la parte central de la distribución, esto mediante la posición relativa de la línea de la mediana dentro de la caja.

4. El tamaño de las colas, lo cual se logra por medio de las longitudes relativas de las líneas exteriores verticales de la caja. La línea más larga determina la dirección de la asimetría.

5. Los valores *outliers:* puntos más allá de las líneas exteriores del rectángulo.

Las particularidades anteriores son más significativas cuando la muestra es grande. Además, estas propiedades son comparables en diferentes conjuntos de datos, lo que hace que las cajas esquemáticas sean útiles especialmente en los estudios comparativos.

A partir de la publicación del libro de Tukey sobre el análisis exploratorio de datos, algunos autores han venido incluyendo este tema en los textos de estadística; por ejemplo, Koopmans (1987), Hoaglin y otros (2000) y Quinn y Keough (2003).

3.5 Análisis inferencial de una muestra aleatoria cuantitativa

La estadística sigue un proceso cíclico. Inicia especificando algunos parámetros de una población, luego selecciona muestras aleatorias y calcula medidas descriptivas; es decir, estadísticas muestrales, que confronta con los modelos teóricos para, finalmente, obtener conclusiones sobre los parámetros poblacionales.

El objetivo de la *inferencia* incluye la obtención de valores para los parámetros, esto es, la estimación puntual y de intervalos, y la comparación de estos valores con los de otras poblaciones, o sea, las pruebas de hipótesis.

Un parámetro es una característica descriptiva de una población, su valor se considera fijo, es desconocido y no calculable.

Un estimador o estadística es una fórmula matemática calculable con los datos de la muestra y su valor, llamado estimación o estimativo, varía de una muestra a otra.

Entonces, estimar un parámetro es hallar un valor numérico, calculado de la muestra, con la esperanza de que dicho valor esté cercano al valor verdadero de la población.

Uno de los propósitos de la Estadística es buscar "buenos estimadores" para los parámetros. Un estimador es "bueno" si cumple con algunas propiedades de tipo teórico como: insesgamiento, eficiencia, suficiencia, resistencia y varianza mínima.

Si la variable de respuesta es continua, el parámetro de posición o localización (tendencia central) más interesante y de mayor uso es la media aritmética μ y el parámetro de escala (variabilidad) es la varianza σ^2. Para variables cualitativas el parámetro más útil es la proporción P.

A continuación se ilustran los conceptos de estimación y de pruebas de hipótesis aplicados a la media poblacional.

1. *Estimación puntual de la media* μ. La media aritmética poblacional μ es desconocida pero puede estimarse, de manera no sesgada, mediante la media aritmética $\bar{y}$ de la muestra. Otros valores promedio como la mediana, la media geométrica, el promedio de extremos $(y_1 + y_n)/2$ o el promedio $(Q_1 + Me + Q_3)/3$,

también estiman la media μ, pero estos estimadores son sesgados, de modo que el mejor estimador es la media aritmética de la muestra. Se escribe: $\hat{\mu} = \bar{y}$, y se lee: un estimador de la media poblacional μ es la media muestral $\bar{y}$. El gorro o techo encima de la letra griega se lee como "un estimador de". De lo anterior, se puede concluir que una buena estimación de la estatura promedio de los estudiantes de la Universidad de Antioquia en el año 1982 era $\hat{\mu} = 1.66$ m.

2. Estimación puntual de la varianza σ^2. El mejor estimador no sesgado de la varianza poblacional σ^2 es la varianza muestral s^2. Se escribe $\hat{\sigma}^2 = s^2$, de aquí se deriva que $\hat{\sigma} = s$; sin embargo, la DE de la muestra no es un estimador óptimo de σ porque no es insesgado. No obstante, la facilidad para calcularla y el hecho que el sesgo sea pequeño, han contribuido a que su uso sea común en los cálculos estadísticos. Otra estimación de la desviación estándar σ podría ser el rango dividido por cinco ($\hat{\sigma} \approx R/5$) pero este es un estimador muy pobre. La varianza y la DE de las estaturas mencionadas en el ejemplo se estiman como $\hat{\sigma}^2 = 0.0124$ y $\hat{\sigma} = 0.11$ m.

3. Estimación de la precisión de un estimador. A la pregunta ¿cuál es el valor del parámetro?, se responde haciendo estimación puntual. Pero, las estimaciones puntuales no informan sobre la cercanía entre el valor calculado y el valor verdadero del parámetro; es decir, no miden la precisión de los estimadores. Entonces, para responder la pregunta ¿qué tan precisa es la estimación?, se requiere una estimación de dicha precisión. La precisión se refiere a la diferencia entre el valor teórico o verdadero y la estimación obtenida de la muestra. Esta diferencia se llama *error de estimación* y se representa por $\delta = |\mu - \bar{y}|$ para el caso de la media aritmética. Como δ es un parámetro, es necesario estimarlo y la mejor estimación es el error estándar de la media $\bar{y}$, esto es, $\hat{\delta} = EE(\bar{y}) = s/\sqrt{n}$. Por tanto, $\hat{\delta}$ es una medida de la precisión del estimador $\bar{y}$. La precisión aumenta cuando decrece $\hat{\delta}$ y lo contrario. El error de estimación de la media de las estaturas es $\hat{\delta} = 0.018$ m.

4. Estimación de un intervalo de confianza para la media μ. El error de la estimación es una magnitud absoluta y su interpretación no es obvia o inmediata con respecto a la variable de estudio. Por esto la estadística propuso evaluar la precisión mediante el análisis de los intervalos de confianza, respondiendo la pregunta ¿qué credibilidad tiene la estimación del parámetro?

La forma natural de expresar *credibilidad* es mediante intervalos matemáticos que incluyan el parámetro con una certeza razonable, o sea alto grado de confiabilidad y este depende del número de veces o frecuencia de aciertos. Así, la confiabilidad se interpreta como una probabilidad de la forma $P(a < \mu < b) = 1 - \alpha$.

El valor más aceptado, entre los investigadores, para la confiabilidad $1 - \alpha$ es el de 0.95 (intervalo del 95% de confianza).

Los límites a y b del intervalo se calculan con la fórmula general:

Estimador $\pm$ *k EE (estimador).*

El valor k depende del grado de confianza asignado al intervalo y de la distribución probabilística del estimador. Para el caso de la media $\bar{y}$, con n grande

puede suponerse que tiene distribución Normal basados en el *Teorema Central del Límite*.

Entonces, un intervalo de confianza $1 - \alpha$ para la media μ tiene la forma $\bar{y} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$, donde $t_{\alpha/2}$ es el valor tabulado de la distribución t de Student con $n - 1$ grados de libertad (véase tabla A.3).

Un intervalo del 95% de confianza para el promedio de las estaturas de los estudiantes es:

$$\bar{y} \pm t_{0.025,39}EE(\bar{y}) = 1.655 \pm 2.02 \times 0.018 = (1.62, 1.69) \text{ m.}$$

La confiabilidad de este intervalo se interpreta como una probabilidad frecuencial, o sea que si se repitiera el muestreo, por ejemplo cien veces en condiciones similares, al menos 95 de los 100 intervalos calculados contendrían el valor de la media que es desconocido. Como no es posible repetir el muestreo cien veces o más, se confía en que el intervalo calculado sea uno de los 95 que contienen el valor de μ.

No se sabe cuál es la estatura promedio de los estudiantes, pero se tiene confianza alta (95 en 100) de que dicho valor se encuentra entre 1.62 y 1.69 m.

La amplitud del intervalo permite razonar sobre la precisión de la estimación. La estimación es más precisa mientras más estrecho sea el intervalo. De todos modos, es el investigador quien puede juzgar el grado de precisión a la luz de sus conocimientos sobre la variable de respuesta.

La precisión y la confiabilidad tienen relación inversa. Si disminuye la confiabilidad, también disminuye la amplitud del intervalo y por tanto aumenta la precisión.

El intervalo del 90% para la media de las estaturas es:

$$\bar{y} \pm t_{0.05,39}EE(\bar{y}) = 1.655 \pm 1.684 \times 0.018 = (1.63, 1.68) \text{ m.}$$

Para confiabilidad del 95%, la amplitud es 0.07 m.

En el caso de confiabilidad del 90%, la amplitud es 0.05 m.

Si se desea aumentar la precisión habría que disminuir la confiabilidad, sin embargo, manejando muestras relativamente grandes y diseños apropiados, es posible mantener tanto la confiabilidad como la precisión en valores científicamente aceptables.

5. Prueba estadística de una hipótesis sobre la media μ. Hasta el momento se ha enfocado el análisis como si el objetivo de la investigación hubiese sido estimar un valor promedio. Supóngase ahora que la muestra fue tomada para comparar las estaturas de los estudiantes de la Universidad de Antioquia con las de los estudiantes de otra universidad colombiana. Las comparaciones son los objetivos específicos más comunes en las investigaciones experimentales. Aquellas se escriben como hipótesis científicas que luego se traducen al vocabulario estadístico en términos de parámetros.

Una *hipótesis científica* es una afirmación acerca de una característica de un hecho de la naturaleza que puede verificarse experimentalmente. Por otra parte, una *hipótesis estadística* es una afirmación acerca de un parámetro en la población que puede probarse o verificarse usando datos o información parcial de la naturaleza.

Las hipótesis científicas deben llevarse al vocabulario estadístico para su verificación. Verificar o probar una hipótesis estadística es concluir sobre su aceptación o rechazo. Por tanto, se trata de un proceso de decisión que consiste en confrontar un resultado experimental con un valor teórico e inferir una respuesta científica.

La estadística trabaja con dos hipótesis: la de investigación, llamada H_1, y su hipótesis complementaria, llamada H_0. El método de prueba es indirecto, esto es, se busca negar o rechazar H_0 para concluir que H_1 es correcta y aceptarla como provisionalmente verdadera.

Entonces, si los datos contradicen H_0, esta se rechaza. Pero si sustentan H_0, esta se acepta. Algunos autores prefieren el concepto "no rechazar" en vez de "aceptar" entendiéndose que el hecho de no rechazar H_0 no implica que haya que aceptarla. La hipótesis H_0 se conoce como hipótesis nula porque la mayoría de las pruebas estadísticas se construyeron de tal manera que, si los datos lo permiten, la conclusión es rechazar H_0. La hipótesis H_1 es una hipótesis alternativa de H_0 y generalmente no se menciona en el proceso de prueba.

En el ejemplo de las estaturas, las hipótesis estadísticas pueden plantearse como sigue, asumiendo arbitrariamente una media de 1.70 para la otra universidad.

H_0: $\mu_0 = 1.70$ m,

H_1: $\mu_0 \neq 1.70$ m (hipótesis bilateral).

En este caso se desea probar que las estaturas de los estudiantes son diferentes en las dos universidades o que la estatura media de los estudiantes de la Universidad de Antioquia puede ser mayor o menor que 1.70 m.

La regla de decisión para rechazar H_0 está basada en los valores $\bar{y}$ y $EE(\bar{y})$ de la muestra. Aquella dice: rechazar H_0 si $\bar{y}$ está por fuera del intervalo $\mu_0 \pm t_{\alpha/2}\, EE(\bar{y})$ y no rechazarla en el caso contrario. Aceptando una probabilidad de error $\alpha = 0.05$, el intervalo es $1.70 \pm 2.02 \times 0.018 = (1.66, 1.74)$. La media 1.655 no pertenece al intervalo calculado, por tanto se rechaza H_0 y se concluye que las estaturas de los estudiantes son diferentes en las dos universidades.

El error del 5% en la prueba anterior significa que se podría estar equivocado al tomar la decisión de rechazar H_0, pero esto ocurrirá en máximo 5 de cada 100 veces que se hiciera un muestreo similar de estudiantes. No puede olvidarse que para usar la distribución t en una prueba sobre la media μ, es requisito que la población de estudio tenga una distribución Normal.

3.6 Significación de una prueba de hipótesis

Para un error α del 5% se encontró, en el ejemplo de las estaturas, que el límite inferior del intervalo de aceptación de H_0 era 1.66 m, la media $\bar{y}$ está por debajo de este límite, por tanto, se rechazó H_0 lo que indica que la probabilidad de error puede disminuirse y aún seguir rechazando esta hipótesis. Pero, ¿cuál es la probabilidad mínima de error que todavía permita rechazar H_0? Dicha probabilidad se llama error crítico de rechazo o *nivel p de significación* de la prueba y su valor exacto ($p = 0.029$)

se obtiene usando programas de computador. La expresión probabilística para calcularlo (prueba bilateral) es de la forma:

$$2P(\text{rechazar } H_0 / H_0 \text{ verdadera}) = 2P(\bar{y} \leq 1.66/\mu_0 = 1.70) = 2P(t_{p,39} \leq -2.27)$$

$$= 2 \times 0.0144 = 0.029.$$

En resumen, cuando se realiza una prueba de hipótesis interesa informar sobre la significancia de la prueba y en vez de decir que se rechaza con $\alpha < 0.05$, es más ilustrativo dar el valor exacto de la probabilidad del error, en este caso, $p = 0.029$. La probabilidad p de una prueba puede interpretarse como un índice de credibilidad de la hipótesis nula en el sentido que mientras menor sea p, mayor es la creencia en la falsedad de H_0.

En los reportes estadísticos son usuales las expresiones siguientes:

1. Si $p > 0.05$, no se rechaza la hipótesis nula y se dice que la *prueba es no significativa.*
2. Si $p < 0.05$, se rechaza la hipótesis nula y se dice que la *prueba es significativa* Además:
   a. Si $0.01 < p < 0.05$, la prueba es estadísticamente significativa.
   b. Si $0.001 < p < 0.01$, la prueba es altamente significativa.
   c. Si $p < 0.001$, la prueba es altísimamente significativa.

Es importante enfatizar sobre el significado del valor de p. La p es la probabilidad de error al rechazar H_0. Mientras más pequeño sea su valor con respecto al valor de referencia ($\alpha = 0.05$), más creíble es la hipótesis de investigación H_1. Cuando p es mayor que 0.05, por ejemplo 0.18, no se rechaza H_0 pero la p sigue siendo una probabilidad de error de rechazo de H_0, es decir, $p = 0.18$ significa que no se rechaza H_0 porque si se rechazara se cometería un error de hasta el 18.0% y en una prueba estadística el error máximo permitido es del 5.0%. Mientras más alto sea el valor de p, más razones hay para "aceptar" la hipótesis nula, o sea que los datos no comprobaron las hipótesis de investigación.

El nivel de significancia de una prueba se escoge de manera subjetiva, pero históricamente la elección ha sido restringida a tres valores: 0.01, 0.05 y 0.10. Uno de los aspectos deleitables de la estadística, como dicen Fisher y McDonald (1978), es que esta tiene lugar entre humanos y se ajusta a sus comportamientos impredecibles. Se conoce el cuento (¿apócrifo?) que dice que cuando R. A. Fisher estaba escribiendo el libro *Statistical Methods for Research Workers,* necesitaba unas tablas que habían sido calculadas en el laboratorio de K. Pearson. Al solicitarle permiso para incluirlas en el libro, Pearson se lo negó, pues existía cierta rivalidad entre ellos. No obstante, Fisher consultó el caso y se le informó que podía usar una parte de la tabla sin violar las leyes sobre derechos de propiedad literaria. Entonces, seleccionó la parte correspondiente a los valores críticos 0.10, 0.05 y 0.01.

La influencia y contribución de Fisher a la estadística fueron tales que hicieron que estos valores se convirtieran en estándares y resultaran aceptados universalmente aún hoy día.

Otra probabilidad de interés en inferencia es la potencia o el poder de una prueba. Cuando se toma una decisión estadística, puede cometerse uno de dos errores.

El error de tipo I, que consiste en rechazar la hipótesis nula siendo esta la hipótesis verdadera. Se representa por α y es el error que usualmente controla el investigador cuando fija su valor, por ejemplo, en el 5%. El otro error, llamado de tipo II, se comete cuando se acepta la hipótesis nula siendo falsa. Se representa por β, pero, para su interpretación, es más fácil trabajar con la probabilidad complementaria $1 - \beta$, llamada *potencia* de la prueba. En símbolos:

$$P(\text{aceptar } H_0 / H_0 \text{ es falsa}) = \beta,$$

$$P(\text{rechazar } H_0 / H_0 \text{ es falsa}) = 1 - \beta = \textit{Potencia}$$

$$= \text{probabilidad de acertar en la decisión de rechazar } H_0.$$

La potencia de una prueba describe la capacidad que tiene la prueba para detectar la falsedad de H_0. Mientras más alta sea la potencia de una prueba, mayor es la seguridad de encontrar una diferencia significativa cuando H_1 es verdadera.

La potencia no es evaluable directamente puesto que es una función de los valores de μ en H_1. Si se da un valor arbitrario a μ en H_1, es decir, fuera del intervalo (1.66, 1.74), por ejemplo, $\mu = 1.64$, la potencia de la prueba sería:

$$\textit{Potencia} = P(\bar{y} \le 1.66 / \mu = 1.64)$$

$$= P(t_{1-\beta} \le 1.31) = 0.9012.$$

Si H_0 fuera falsa, esta prueba la rechazaría en aproximadamente el 90% de las veces que se realizara. La potencia depende del valor elegido para μ. Si $\mu = 1.67$, la potencia sería del 36.2% únicamente.

3.7 Pruebas de bondad de ajuste a la distribución Normal

La inferencia estadística paramétrica supone que las muestras provienen de poblaciones con distribuciones conocidas. Por otro lado, se ha encontrado en muchas ramas de la ciencia que las medidas experimentales repetidas se ajustan al modelo Normal, al menos en forma aproximada: mediciones de variables como peso, longitud, volumen, concentración, densidad óptica, temperatura, presión sanguínea y muchas más, se analizan estadísticamente bajo la suposición de normalidad de la población de la cual se extrajo la muestra. No obstante, se recomienda asegurarse de que la suposición es válida para la muestra que se tiene; para ello se dispone de un conjunto grande de pruebas, cada una con sus características peculiares.

Las primeras pruebas de ajuste al modelo Normal fueron la prueba *ji*-cuadrado de Pearson y la prueba de Kolmogorov-Smirnov. La primera fue diseñada para datos categóricos, de modo que al aplicarla a datos continuos, estos deben agruparse en intervalos. Tiene como desventaja que requiere un número grande de datos (100 o más) y además es muy sensible tanto al número de clases como a las frecuencias esperadas en los intervalos de los extremos. Detalles de esta prueba pueden verse en Díaz y Gutiérrez (1995). La prueba de Kolmogorov-Smirnov, por el contrario, fue

diseñada para muestras de poblaciones continuas, no requiere agrupar los datos y puede aplicarse a muestras virtualmente de cualquier tamaño.

En ambas pruebas deben estimarse μ y σ mediante $\bar{y}$ y s, convirtiéndose así en pruebas asintóticas o aproximadas. Su aplicación se encuentra en varios textos como Gibbons (1993), Steel y Torrie (1997) y Zar (1999); se remite al lector a estas fuentes con el fin de dar espacio a otra prueba más recomendada actualmente, sobre todo cuando se tienen pocos datos.

Las pruebas de normalidad tienen como hipótesis:

H_0: la muestra proviene de una población con distribución Normal.

H_1: la muestra no proviene de una población con distribución Normal.

La regla de decisión es diferente para cada prueba, lo mismo que las estadísticas de comparación y las tablas requeridas para tomar la decisión.

Prueba de normalidad de Shapiro-Wilk

Esta es una de las pruebas con mayor sensibilidad a la no normalidad, en esta no es necesario calcular la media ni la varianza de la muestra para incluirlas en la hipótesis, pero requiere dos tipos de tablas para su aplicación, Shapiro y Wilk (1965). Los autores han proporcionado tablas para $n \leq 50$, las cuales han sido reproducidas en Anderson y McLean (1974) y en Gill (1978). Shapiro y Francia (1972) propusieron una versión más simplificada de esta prueba.

Ejemplo 3.5 Serna (1977), midió la longitud total corporal (mm) en una muestra de 11 mayos capturados en Medellín (Colombia) y sus alrededores. Usando la prueba de Shapiro-Wilk, comprobar si las longitudes totales de dichas aves siguen una distribución Normal.

La prueba consta de los siguientes pasos:

a. Se ordenan los datos de menor a mayor: $y_1 \leq y_2 \leq \ldots \leq y_n$.

165, 223, 225, 231, 235, 235, 237, 240, 245, 245, 245

b. Se calcula la suma de cuadrados: $SC = \sum_i (y_i - \bar{y})^2$.

$$SC = \sum_i y_i^2 - (1/n)\left(\sum_i y_i\right)^2 = 5172.545.$$

c. Si n es par, $n = 2k$, y se calcula $b = \sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i)$.

Si n es impar, $n = 2k + 1$, se omite la mediana y se calcula b como en el caso anterior. Se tiene que $n = 11$, luego se omite la $Me = 235$, y con $2k + 1 = 11$, resulta $k = 5$. Entonces, $b = a_{11}(y_{11} - y_1) + a_{10}(y_{10} - y_2) + a_9(y_9 - y_3) + a_8(y_8 - y_4) + a_7(y_7 - y_5)$.

Los coeficientes a_j se leen en la tabla A.4, resultando $b = 0.5601(245 - 165) + 0.3315(245 - 223) + 0.2260(245 - 225) + 0.1429(240 - 231) + 0.0695(237 - 235) = 58.046$.

d. Se calcula la estadística de comparación $W = b^2 / SC = (58.046)^2 / 5172.545 = 0.651$.

e. La regla de decisión dice: si $W < W_{\alpha,n}$ se rechaza H_0. Valores pequeños de W indican falta de normalidad en la población.

Para $\alpha = 0.05$, se lee en la tabla A.5 que $W_{0.05,11} = 0.850$. Entonces, con error del 5%, se concluye que los datos no provienen de una población Normal. No obstante, mirando

la muestra se encuentra que el primer valor es muy pequeño con respecto a los demás. En efecto, este dato corresponde a un pichón, según el autor. El dato 165 se tratará en la sección siguiente como un *outlier.*

Se han propuesto en la literatura estadística varias pruebas de bondad de ajuste a distribuciones específicas, en particular a la Normal. Por ejemplo, Anderson y Darling (1952), D'Agostino (1971) y Thode Jr. (2002).

Determinación gráfica de normalidad

Un método diferente para examinar si un conjunto de datos se ajusta al modelo Normal, consiste en graficarlos en una escala probabilística normal. Los datos se grafican localizando en la abscisa la escala real de los datos y en la ordenada la frecuencia relativa acumulada. Si los puntos forman una línea recta, puede concluirse que los datos vienen de una población con distribución Normal.

El problema que se presenta con este método es que la linealidad del gráfico debe apreciarse subjetivamente. Resulta un método rápido y fácil para juzgar la no normalidad de un conjunto de datos, sobre todo cuando es posible hacer el gráfico en el computador (véanse figuras 3.6 a 3.8).

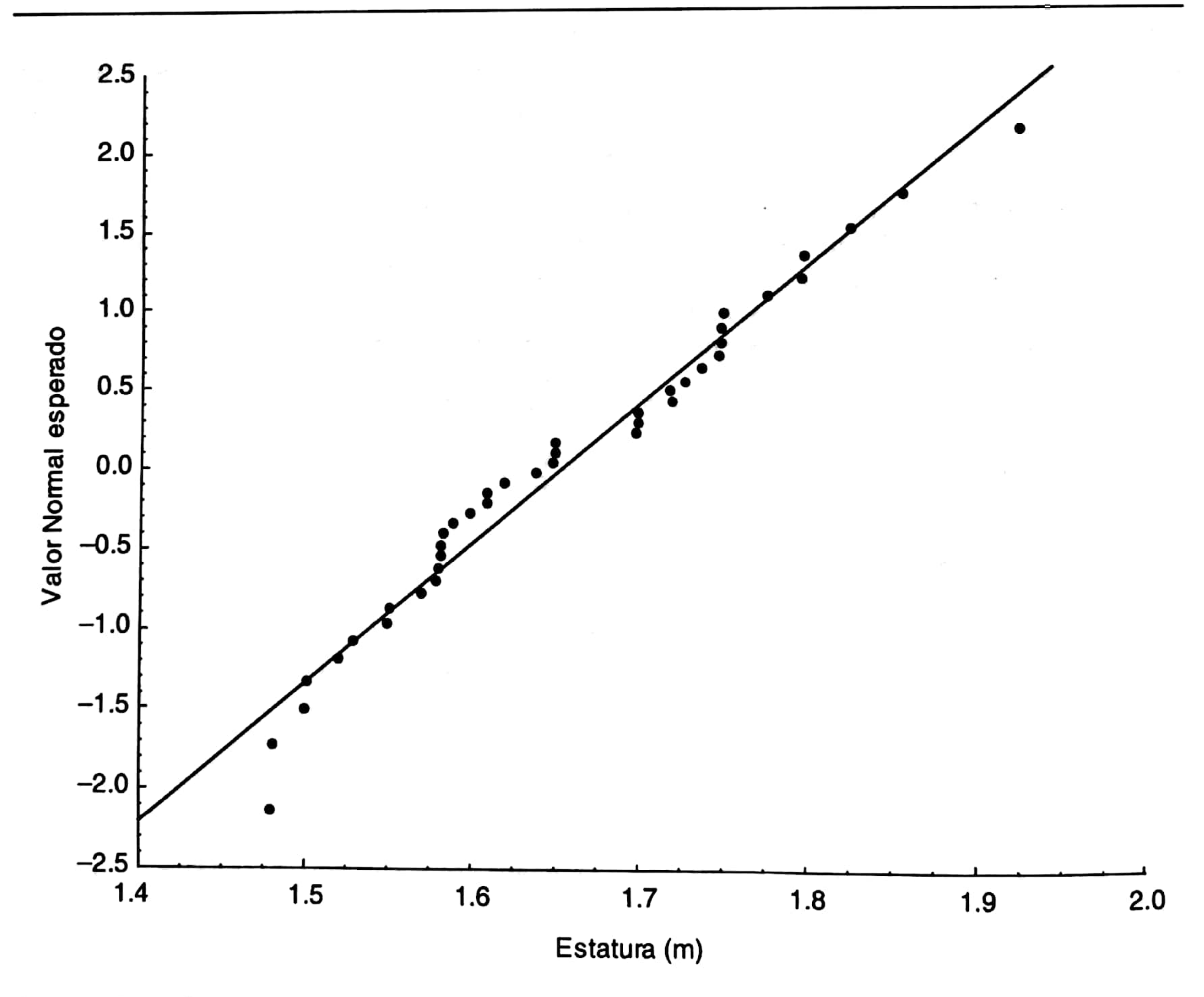


Figura 3.6 Gráfico de probabilidad Normal para el ejemplo 3.1

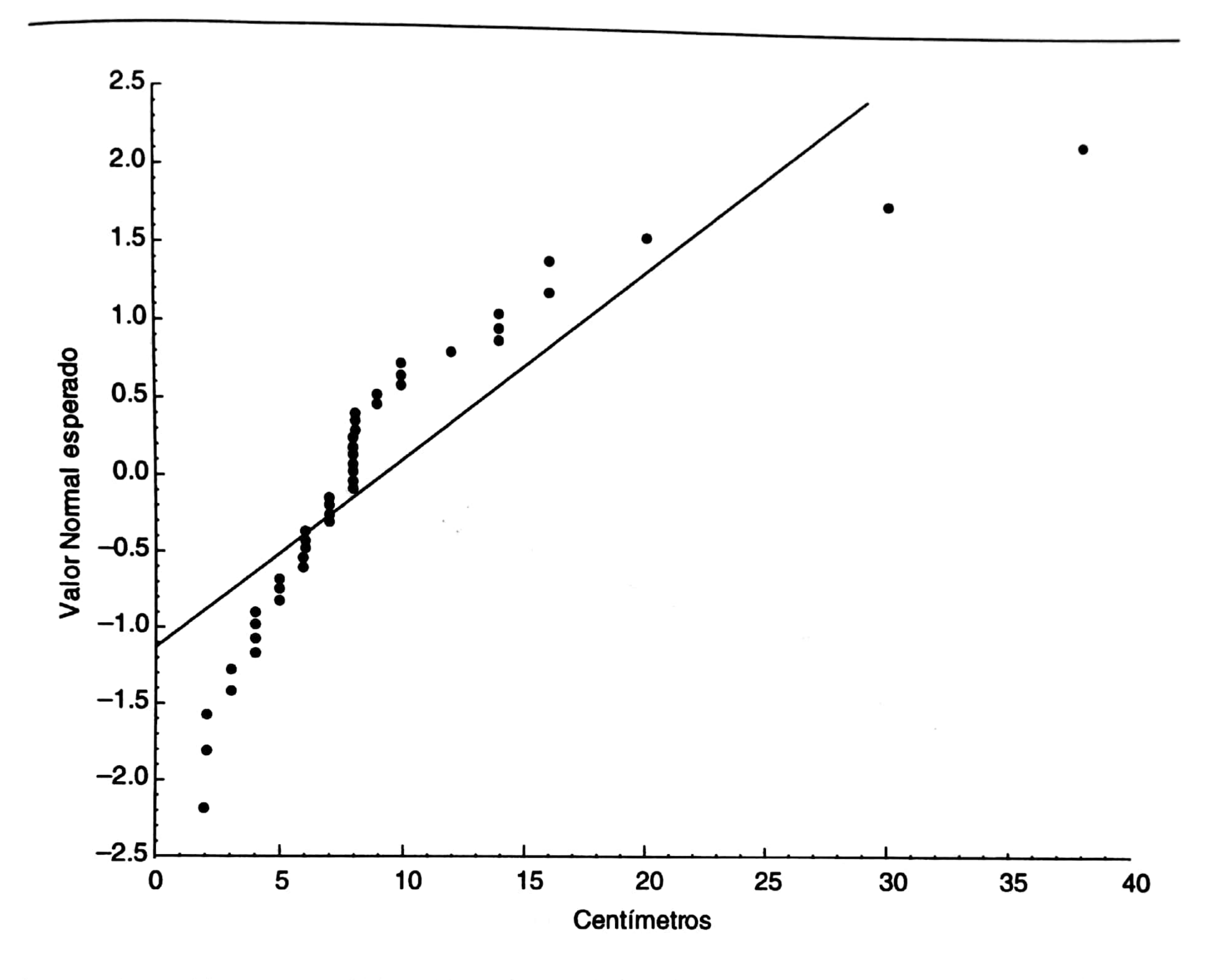


Figura 3.7 Gráfico de probabilidad Normal para el ejemplo 3.4

El método gráfico tiene como principio teórico que si los datos tienen distribución Normal, la figura correspondiente es una línea recta. En la práctica se debe razonar en forma inversa: si los puntos están lejos de formar una recta, entonces no hay normalidad en los datos. La recta es una situación ideal que no se logra con una muestra, pero si el gráfico es aproximadamente lineal, o sea, los puntos van zigzagueantes alrededor de una línea recta, se puede suponer normalidad.

En la figura 3.6, correspondiente a los datos de las estaturas de los estudiantes, se observa buen ajuste a la Normal para dicha muestra. Si se presenta una curvatura pronunciada, se concluye que hay asimetría en los datos. En la figura 3.7 aparece el tipo de curvatura que indica asimetría positiva; en efecto, corresponde a los datos del ejemplo 3.4 ya analizados. La figura 3.8 es representativa de una muestra con asimetría negativa.

El objetivo de una prueba de normalidad no es tanto establecer normalidad sino evaluar el alejamiento de los datos respecto a un modelo Normal. En este sentido, el método gráfico es una excelente ayuda para tomar la decisión apropiada, pues el ojo humano detecta más fácilmente los puntos alejados con respecto a una recta que con respecto a otra figura.

En vez de la escala de probabilidad Normal pueden usarse otras escalas en el

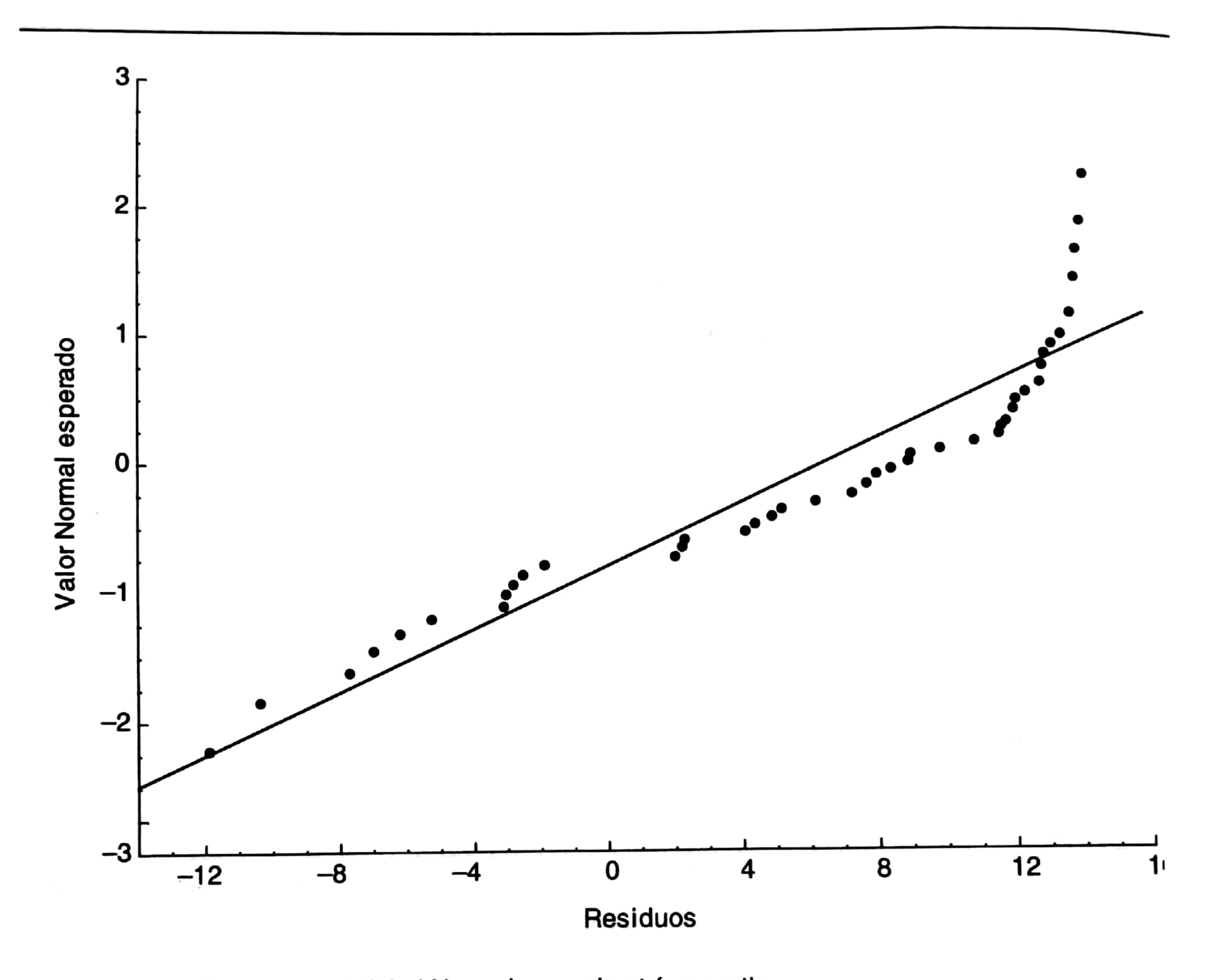


Figura 3.8 Gráfico de probabilidad Normal con asimetría negativa

eje vertical. El método *probit*, por ejemplo, transforma la escala de porcentajes acumulados a una escala del 2.0 al 9.0; esta escala es apropiada para muestras grandes ($n > 100$) y se aplica a los datos agrupados. La escala *rankit*, por el contrario, varía aproximadamente entre −2.0 y +2.0 para tamaños de muestra menores que 40, es simétrica y responde adecuadamente aun para muestras entre 10 y 15 datos. Los valores probit y rankit se leen directamente en tablas como las que trae Wardlaw (2000). Si los puntos graficados en estas escalas forman una línea recta, se concluye que la transformación normaliza los datos.

3.8 Datos *outliers*

Los investigadores que trabajan con resultados obtenidos de mediciones u observaciones tomadas bajo las mismas condiciones encuentran, frecuentemente, que entre sus datos hay uno, dos o más que son diferentes al resto. Estos datos se llaman *outliers*, voz inglesa usada por los estadísticos a falta de una traducción unánime de dicha palabra. Datos *outliers* son valores extraños, atípicos o extremos que están tan alejados de los valores centrales de la muestra que su presencia es dudosa y son candidatos para ser excluidos de los análisis estadísticos.

El primer problema del investigador es decidir si el dato *outlier* se excluye en el momento de hacer cálculos. Aunque el dato no se tenga en cuenta en los cálculos, deberá quedar registrado y, en lo posible, con explicaciones de las causas que lo originaron. Algunos datos resultan dudosos a causa de, por ejemplo, recipientes sucios, cambios bruscos en las condiciones ambientales, registros equivocados o mal funcionamiento de los aparatos, en estos casos la exclusión del dato resulta justificada. Cuando nada explica el dato, su exclusión puede ser asunto del sentido común. La experiencia del investigador en el manejo de sus datos le permite, muchas veces, reconocer aquellos que deben descartarse de los cálculos.

La discusión anterior tiene importancia puesto que la exclusión o inclusión de datos extremos afecta, tanto la precisión de las estimaciones como la sensibilidad de las pruebas. Algunos investigadores sostienen que es mejor perder un dato posiblemente *bueno* que ganar un dato *malo*, y obtener resultados sesgados y viciados.

Existen varios criterios estadísticos para reconocer o catalogar un dato como *outlier* y para aceptarlo o rechazarlo en los cálculos. Los gráficos de ramas y de cajas esquemáticas permiten identificar datos sospechosos. Además, existen algunas pruebas para determinar si estos pueden considerarse homogéneos con el resto de la muestra, Dixon (1950), Grubbs (1969).

Algunas pruebas de normalidad son también, indirectamente, pruebas de detección de valores *outliers*, en cuanto a que la no normalidad puede deberse a la presencia de uno o más datos extremos en la muestra. Excluidos estos datos de la muestra, pero no de la investigación, la prueba puede indicar normalidad; este es el caso que se presenta en el ejemplo 3.5. En efecto, suprimiendo el dato 165 y aplicando la prueba de Shapiro y Wilk al resto de la muestra, se obtiene:

$$b = 0.5739(245-223) + 0.3291(245-225) + 0.2141(245-231) + 0.1224(240-235) + 0.0399(237-235).$$

$$b = 22.897.$$

$$SC = 576.900.$$

$$W = 524.273 / 576.900 = 0.9088 \text{ y } W_{0.10,10} = 0.869.$$

Se acepta normalidad de los datos con $p > 0.10$.

También existen pruebas para detectar simultáneamente varios datos *outliers*. Tietjen y Moore (1972) proponen dos pruebas que requieren múltiples tablas para obtener sus valores críticos, pero que poseen buena potencia. Un método más simple, aunque menos potente, es el procedimiento secuencial: se fija un nivel de error α y se prueba normalidad para la muestra, si la no normalidad se debe a un dato sospechoso, este se excluye y se repite la prueba para la muestra reducida; el proceso continúa hasta que se encuentre un subconjunto que no contenga datos *outliers*.

Una desventaja del método secuencial es que resulta insensible a efectos de *enmascaramiento*, o sea, la inhabilidad de la prueba para identificar datos *outliers* cuando aparece un subgrupo de estos que están más cercanos entre sí que del resto de la muestra. El ejemplo siguiente ilustra el procedimiento secuencial mediante la prueba de Shapiro y Wilk.

Ejemplo 3.6 La medición de fosfatos en la superficie del agua en el embalse de La Fe, Ramírez y Machado (1982), durante 13 meses —enero 1978 a febrero 1979— dio los siguientes resultados:

2.30, 1.65, 1.30, 0.90, 1.10, 1.30, 1.12, 1.25, 2.00, 4.60, 6.00, 1.12, 3.95.

El dato 6.00 es muy alto comparado con el resto. Se aplica la prueba de normalidad a los 13 datos y se encuentra:

$$b^2 = 23.8354, SC = 31.3703 \text{ y } W = 0.760 < 0.866 = W_{0.05,13}.$$

Se rechaza el valor 6.00 y se prueba normalidad para los 12 datos restantes:

$$b^2 = 11.7427, SC = 15.7206 \text{ y } W = 0.747 < 0.859 = W_{0.05,12}.$$

Se rechaza el dato 4.60 y se repite para $n = 11$:

$$b^2 = 5.6844, SC = 7.6648 \text{ y } W = 0.742 < 0.850 = W_{0.05,11}.$$

Se rechaza el valor extremo 3.95 y se aplica la prueba a los 10 datos que quedan:

$$b^2 = 1.5305, SC = 1.7716 \text{ y } W = 0.864 > 0.842 = W_{0.05,10}.$$

Se concluye entonces, que dicha muestra tiene tres valores extremos que deben excluirse si se desea manejarla como proveniente de una población con distribución Normal. La figura 3.9 representa la muestra gráficamente, allí se observa el sesgo positivo causado por los tres datos extremos.

Cuando la muestra es pequeña, el rechazo de un dato extremo hace que la media y la varianza cambien de tamaño y, por tanto, deberá tenerse suficiente información para decidir el rechazo. No se recomienda, por ejemplo, la práctica de hacer tres mediciones y desechar la más alejada de las otras dos, aunque la costumbre indique que se haga este rechazo. Es más confiable tomar la mediana de las tres lecturas, que la media aritmética de las dos que quedan después de desechar la más extrema.

El manejo de datos *outliers* en las muestras ha sido estudiado ampliamente en la literatura estadística. Uno de los textos más completo sobre este tema es el de Barnett y Lewis (1994).

3.9 Tamaño de la muestra

Uno de los objetivos de las investigaciones experimentales es la comparación de medias poblacionales entre sí o con respecto a valores de referencia conocidos. En el diseño del experimento surge la pregunta sobre el tamaño de la muestra o el número requerido de repeticiones para realizar una inferencia estadística con buena

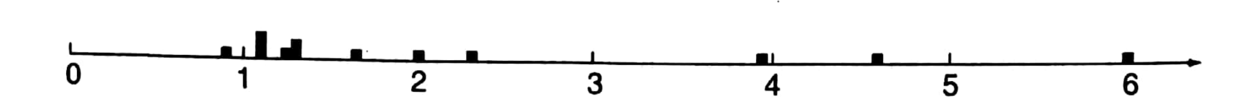


Figura 3.9 Datos del ejemplo 3.6

precisión. La primera pregunta debe referirse al tipo de inferencia que se planea efectuar a partir de esa muestra.

Algunos métodos estadísticos proporcionan fórmulas para estimar tamaños de muestra, ya sea en estimaciones de intervalos de confianza o en pruebas de hipótesis. Considérese el caso más simple, o sea, comparar una media poblacional μ con un valor conocido μ_0 mediante una sola muestra. Su tamaño está relacionado con las respuestas a las siguientes preguntas:

1. ¿Cuál es la mínima diferencia significativa que se desea detectar entre las dos medias? Debe establecerse la diferencia $\delta = |\mu - \mu_0|$, o sea, la diferencia científicamente significativa según la apreciación del investigador. Esta diferencia es un valor hipotético llamado *tamaño del efecto,* así que la pregunta podría plantearse como: ¿cuál es el tamaño del efecto que se desea considerar en la prueba estadística?
2. ¿Con qué probabilidad se espera obtener una diferencia igual o menor que δ? Elegir las probabilidades α y β.
3. ¿Cuál es la variabilidad de la población? ¿Se conoce σ?

La respuesta a la segunda pregunta consiste en fijar los errores de tipo I (falsos positivos) y de tipo II (falsos negativos) que pueden cometerse al inferir, de esa muestra, una decisión equivocada. Es claro que el experimentador no comete ambos errores en una sola decisión. Si acepta la hipótesis alternativa, puede caer en el error de tipo I cuando la hipótesis correcta sea H_0; y si acepta la hipótesis nula cuando la hipótesis correcta es H_1, cae en el error de tipo II únicamente.

La escogencia de las probabilidades de error es casi siempre subjetiva, pero debe tenerse en cuenta que si se eligen muy pequeñas, el tamaño de la muestra exigida será grande. La gravedad de cada uno de los errores al tomar las decisiones, debe ser una guía en esta elección. Los valores típicos que se encuentran en la literatura son $\alpha = 0.05$ y $\beta = 0.10$ o 0.20.

Si la varianza poblacional es conocida, la fórmula para el tamaño muestral es:

$$n = (Z_\alpha + Z_\beta)^2 \, \sigma^2 / \delta^2 \text{ (prueba unilateral)}, \tag{3.1}$$

Los valores de Z se leen en la tabla A.2 de la distribución Normal. Para pruebas bilaterales debe usarse $Z_{\alpha/2}$ en vez de Z_α, pero Z_β no cambia.

Ejemplo 3.7 Calcular el tamaño de la muestra para comparar la estatura promedio de los estudiantes de la Universidad de Antioquia con respecto a la media de otra población.

Para efectos de comparación se establece, en forma arbitraria, que es significativa una diferencia $\delta = 0.03$ m entre medias. Se supone que $\sigma^2 = 0.0124$ es el valor poblacional, además, para $\alpha = 0.05$, $Z_{\alpha/2} = 1.96$ y para $\beta = 0.10$, $Z_\beta = 1.28$. Entonces:

$$n = (1.96 + 1.28)^2 \times 0.0124 / 0.0009 = 144.6$$

Se requiere entrevistar 145 estudiantes para garantizar que la prueba estadística no sobrepase los riesgos establecidos inicialmente.

No siempre el experimentador dispone de los recursos necesarios para obtener una muestra grande. Supóngase que solo es posible entrevistar 80 estudiantes. Dejando los mismos valores para α, δ y σ, se puede calcular el error β y decidir si con esa muestra la prueba es todavía eficiente:

$80 = (1.96 + Z_\beta)^2 \times 0.0124 / 0.0009,$

$Z_\beta = 0.45,$

$\beta = 0.33.$

La potencia de la prueba sería del 67% o sea que, si la diferencia de 0.03 m es real, la prueba estadística la detectaría en el 67% de las veces. En el otro sentido, significa que el 33% de las veces la prueba indica que no hay diferencias entre las estaturas cuando en realidad sí las hay. A menos que el investigador acepte un error de este tamaño, el estudio realizado con 80 datos deberá considerarse como un estudio piloto sin resultados concluyentes. Su objetivo sería exploratorio más bien que inferencial.

El ejemplo anterior ilustra el razonamiento y los cálculos para hallar el tamaño de la muestra en el caso de la comparación de una media; en realidad no es un caso típico experimental, pero la metodología es completamente similar cuando se trata de diseñar una situación que requiere experimentación.

La varianza poblacional es con frecuencia desconocida, lo cual hace que la fórmula anterior no sea calculable. En este caso, debe estimarse σ^2, ya sea de una muestra preliminar, de experimentos anteriores o de algún tipo de conjetura razonable. Si la estimación es $\hat{\sigma}^2$, entonces:

$$n_t = (t_\alpha + t_\beta)^2\, \hat{\sigma}^2 / \delta^2. \qquad (3.2)$$

Cuando no se conocen los *gl* para *t*, debe calcularse primero un valor *n* usando la distribución *Z*, y tomar este *n* para calcular *t*. Si el tamaño muestral es pequeño, su cálculo requiere un proceso iterativo que no se ilustrará aquí, pero se recomienda su lectura especialmente en el libro de Diamond (2001). También Desu y Raghavarao (2004), Cox (1992) y Natrella (2005) discuten ampliamente el problema del tamaño muestral para diferentes diseños comparativos. Chow, Shao y Wang (2003), trae tablas y fórmulas para determinar el tamaño *n* en diferentes condiciones clínicas.

Se puede, entonces, concluir que la respuesta sobre el tamaño muestral no es ni simple ni única. La fórmula para hallar *n* es una conexión entre el pensamiento estadístico y el campo científico donde se desea investigar. Los errores α y β reflejan las leyes de la estadística y la distribución Normal, el tamaño del efecto δ refleja los requisitos de la práctica científica en el experimento, y σ^2 representa la variabilidad de los fenómenos de la Naturaleza.

Mirando estos factores en la fórmula (3.2), se concluye que el tamaño del experimento aumenta cuando los riesgos de error son más pequeños. Por otra parte, si se logra disminuir la varianza, también disminuye el tamaño muestral. Finalmente, mientras más pequeño sea el tamaño del efecto o la diferencia que se desea detectar, mayor será el tamaño de la muestra requerida en la prueba.

Si se desea estimar un intervalo de confianza del $(1 - \alpha)100\%$ para la media μ, el tamaño de la muestra será:

$$n = \left(t_{\alpha/2}\, \hat{\sigma}/\delta\right)^2,$$

donde δ representa la precisión deseada por el investigador.

La estimación de tamaño de la muestra es aún un tema de interés tanto para

los estadísticos como para los investigadores. Además de la literatura especializada según el objetivo y el tipo de investigación, se puede elegir entre una variedad, cada día creciente, de programas de computador sobre determinación de tamaños muestrales. Esta alternativa es recomendada cuando se quiere escoger entre varios tamaños u opciones para armonizar las restricciones de economía y precisión.

En investigaciones con humanos y animales, el tamaño de la muestra, fuera del factor económico, pasa a contemplar el aspecto ético. En un experimento sobredimensionado se expone un número innecesario de sujetos a tratamientos potencialmente peligrosos, mientras que en uno infradimensionado, algunos sujetos son expuestos a un sufrimiento que no conduciría al avance del conocimiento.

Por otro lado, hay situaciones en las que el número de unidades de estudio está prefijado ya sea por recursos financieros, por tiempo o por orden de autoridad superior.

Es necesario, entonces, enfocar el problema hacia otros aspectos de la calidad del estudio, por ejemplo, el tamaño del efecto (precisión) según el objetivo científico, la potencia de las pruebas estadísticas y el diseño experimental elegido.

El análisis del poder o potencia de una prueba estadística ha ganado popularidad en la literatura biológica, Thomas y Juanes (1996). La potencia está relacionada con las pruebas de hipótesis y, por tanto, con el tamaño de la muestra y el tamaño del efecto. La escogencia de un tamaño de efecto significativo para el diseño experimental es un aspecto crucial en el análisis de la potencia, sobre todo por que existen diferentes medidas del tamaño del efecto para una prueba estadística dada.

Los análisis de la potencia pueden ser prospectivos cuando se realizan paralelos al diseño de la investigación, estos permiten explorar el alcance de los objetivos del estudio mediante la variación en los componentes de la potencia (tamaño muestral, α, tamaño del efecto y varianza poblacional).

Estos análisis son retrospectivos cuando se realizan después de haber terminado la investigación y analizado los datos. Para algunos investigadores el concepto de potencia retrospectiva no es válido en el marco de las pruebas de hipótesis, además, sostienen que la estimación de la potencia retrospectiva es raramente útil y debería reemplazarse por un intervalo de confianza para el tamaño del efecto, Scheiner y Gurevitch (2001).

Ejercicios

3.1 De un cultivo de botón de oro *(Ranunculus repens)*, se tomó como muestra 35 flores maduras y se contó el número de estambres por flor. Los resultados fueron:

60	54	55	70	64	53	43	53	47	58	51	67	50	42	39	38	46	40
37	35	22	64	57	75	73	78	52	45	50	62	55	58	57	55	49	

a. Construir el gráfico de ramas para estos datos y analizarlo.
b. Calcular las medidas descriptivas fundamentales.
c. Construir la caja esquemática y analizarla.
d. Construir un intervalo de confianza del 95% para el promedio de estambres por flor.

e. Probar la hipótesis de que el número promedio de estambres por flor es igual a 65. Dar el valor p de esta prueba.
f. Estimar el tamaño de la muestra en un estudio futuro si se acepta que una diferencia de cinco estambres es significativa.
g. Probar la normalidad de estos datos.

3.2. En un estudio sobre el rendimiento profesional de las mujeres, se consideró como una de las variables el coeficiente de testosterona libre CTL en la sangre (medida en mg/100 mL). Los datos obtenidos fueron:

1.2	2.9	3.4	5.2	1.5	3.0	3.5	5.3	2.1	3.1	3.6	5.7	
2.2	3.2	3.6	6.1	2.2	3.3	4.8	6.2	2.7	7.4	5.9	7.8	4.2

a. Calcular las medidas descriptivas fundamentales para esta muestra.
b. Construir el gráfico de ramas y el de la caja esquemática y analizarlos.
c. Probar la normalidad para estos datos. Dar el valor p de la prueba.
d. Estimar el CTL promedio para mujeres profesionales.
e. Probar la hipótesis de que el CTL promedio de las mujeres profesionales es mayor que el de las mujeres que trabajan en las labores del hogar ($\mu = 2.4$).
f. Calcular el tamaño de la muestra, con $\delta = 1.0$, para probar la hipótesis anterior.
g. Analizar el tamaño de muestra $n = 30$ para un estudio similar al planteado anteriormente. ¿Cuál es el valor de β en este caso?; ¿cuál es la potencia de la prueba?

3.3. El pH del agua en el embalse de El Peñol, departamento de Antioquia (Colombia), medido mensualmente entre marzo de 1973 y marzo de 1974 fue:

7.35	7.33	7.08	7.35	7.27	6.92	6.92	6.97	6.95	7.12	7.10	7.15	6.96

a. Construir un gráfico sobre la recta real para esta muestra.
b. Analizar la normalidad de estos datos.
c. Aplicar el método secuencial, si es procedente, y discutir valores extremos dudosos.

3.4. Los siguientes datos representan sulfuro de hidrógeno, producido en la fermentación anaeróbica de aguas negras luego de 42 horas a 37 °C en nueve series.

210	221	218	228	220	227	223	224	192

a. Probar normalidad para estos datos.
b. ¿Podría catalogarse el dato 192 como un dato *outlier*?

3.5. Discutir la respuesta a las siguientes preguntas:

a. Tipificar la variable de respuesta en los ejercicios 3.1 y 3.4.¿Puede la variable discreta (ejercicio 3.1) analizarse como si fuera continua?

b. ¿Pueden reemplazarse los histogramas por cajas esquemáticas en una publicación científica?
c. ¿Podría una prueba de hipótesis estadística asimilarse a un intervalo de confianza construido con los mismos datos?
d. ¿Es la normalidad de una población un hecho real o meramente un requisito teórico?
e. ¿Si dos investigadores sobre el mismo tema manejan diferentes tamaños muestrales, serán los resultados esperados igualmente diferentes? Entonces, ¿qué influencia tiene el tamaño muestral?

4

Diseños completamente aleatorizados

4.1 Diseño completamente aleatorizado

El diseño más simple y general es el diseño completamente aleatorizado, DCA. Este es un diseño sin restricciones de aleatoriedad, lo que quiere decir que los tratamientos se asignan a las unidades experimentales sin formar subgrupos, como sería en el caso del diseño de bloques. El ejemplo siguiente es un problema típico para ilustrar el diseño completamente aleatorizado.

Ejemplo 4.1 Se tienen cuatro tratamientos médicos T_1, T_2, T_3 y T_4, aplicables a la misma enfermedad, y se desea comparar su efectividad. La población es el conjunto de pacientes que llegan a un hospital y que se les diagnostica esa enfermedad. El factor de análisis o variable independiente es la terapia que se va a emplear y tiene cuatro niveles o tratamientos.

Una forma de medir la efectividad del tratamiento es mediante el número de días que permanece el paciente en el hospital hasta su recuperación total. Este número de días es la variable de respuesta que se obtiene de cada paciente tratado, y los pacientes con el diagnóstico de la enfermedad, son las correspondientes unidades experimentales. Para este estudio se seleccionaron veinte pacientes mediante la técnica del muestreo, la cual dice cómo seleccionar los pacientes si se conoce cierta información sobre la conformación de la población —acceso de los pacientes al hospital, tiempo que dura la investigación, etc.—. De esos veinte pacientes, se asignaron cinco a cada tratamiento.

Para que el diseño sea completamente aleatorizado, se requiere, en primer lugar, seleccionar una muestra al azar de veinte pacientes de la población y, en segundo lugar, distribuir los tratamientos a los pacientes también en forma aleatoria —estos procesos pueden llevarse a cabo de muchas maneras.

Si la llegada de pacientes es frecuente, los pacientes que se requieren podrán seleccionarse entre los primeros 200 o 500 que ingresan, lo cual se logra utilizando una tabla de números aleatorios. Si la disponibilidad de pacientes es menor, podría seguirse un muestreo sistemático, por ejemplo, uno de cada dos o uno de cada tres que llegan.

Una vez obtenida la muestra, se numeran los pacientes, o unidades experimentales, en el orden de selección como $u_1, u_2, \ldots, u_{20}$, para asignarles los respectivos tratamientos. Esta asignación debe ser aleatoria y una de las muchas formas posibles de hacerlo es la siguiente:

1. Se numeran los tratamientos como T_1, T_2, T_3 y T_4.
2. Se asigna un tratamiento a cada unidad experimental mediante números aleatorios obtenidos, por ejemplo, de la tabla A.1. Así, los números 1 y 5 corresponden a T_1, los números 2 y 6 a T_2, el 3 y el 7 a T_3, el 4 y el 8 a T_4, y se desechan el 0 y el 9.

3. Se tiene el siguiente orden, y la asignación correspondiente (véase tabla 4.1), partiendo de la primera cifra de la columna 7 de la tabla A.1. Por ejemplo, la unidad u_1 va al tratamiento T_2, porque el primer número aleatorio fue el 6, el cual corresponde a dicho tratamiento, y así para las demás unidades.

Orden aleatorio: 6 2 6 3 8 6 5 4 8 3 5 5 3 6 3 3 1 1
Orden de las unidades: u_1 u_2 u_3 u_4 . . .

4. Se observa que cuando un tratamiento ha recibido todas las unidades, es ignorado al continuar con el proceso. La correspondencia anterior no se aplicaría en el caso de grupos desiguales.

El DCA es flexible; se ajusta a cualquier número de tratamientos, cada uno con igual o diferente número de repeticiones. En particular, las observaciones perdidas no crean dificultades en el análisis de este diseño. La principal desventaja es que exige una muestra homogénea, la cual es difícil de obtener cuando hay muchos tratamientos, muchas repeticiones o ambos.

Realizado el experimento, se tiene el número de días que permaneció hospitalizado el paciente, y los datos pueden organizarse como en la tabla 4.2.

A partir de esta tabla se inicia el análisis estadístico de los datos, tarea que se abordará en la sección 4.3.

4.2 Principios del análisis de varianza

Scheffé (1999) trae, en el primer capítulo de su libro, la siguiente definición, que él mismo considera tentativa: "El análisis de varianza es una técnica estadística para analizar mediciones que dependen de varias clases de efectos que operan simultáneamente, para estimar los efectos y para decidir cuáles efectos son importantes". Las mediciones u observaciones pueden hacerse en una ciencia experimental como la genética, o en una no experimental como la astronomía. Una teoría para analizar mediciones tiene, naturalmente, implicaciones importantes sobre la manera como deberá planearse

Tabla 4.1 Asignación aleatoria de cuatro tratamientos a 20 unidades experimentales

Repeticiones	T_1	T_2	T_3	T_4
1	u_7	u_1	u_4	u_5
2	u_{11}	u_2	u_{10}	u_8
3	u_{12}	u_3	u_{13}	u_9
4	u_{17}	u_6	u_{15}	u_{19}
5	u_{18}	u_{14}	u_{16}	u_{20}

Tabla 4.2 Días de permanencia en el hospital por paciente y por tratamiento

Repeticiones	T_1	T_2	T_3	T_4	
1	41	48	40	40	
2	44	49	50	39	
3	45	49	44	46	
4	43	49	48	46	
5	42	45	50	41	
Totales	215.0	240.0	232.0	212.0	899.0
Medias	43.0	48.0	46.4	42.4	44.95
Varianzas	2.5	3.0	18.8	11.3	

el experimento o hacerse las observaciones, es decir, sobre el diseño experimental. Para G. W. Snedecor, el análisis de varianza es "una técnica para segregar de grupos de datos comparables la variación atribuible a fuentes específicas"; citado por Folks (1981).

La base teórica del análisis de varianza fue desarrollada y publicada por R. A. Fisher (1923). Cinco años antes, este matemático-estadístico inglés (1890,1962) había usado los términos varianza y análisis de varianza no para estudiar la varianza propiamente sino la variación experimental. Hoy día, el análisis de varianza es una técnica estadística que permite analizar datos provenientes de un experimento aleatorio comparativo.

El análisis de varianza, en adelante ANOVA, es un conjunto de procedimientos que se ajustan a las características del diseño experimental usado en la obtención de los datos. No es un método fijo, por el contrario, se adapta a cada diseño y se puede clasificar de distintas maneras.

Si el diseño es unifactorial, se aplica el ANOVA de clasificación simple o de una vía. Si es multifactorial, el ANOVA correspondiente será de dos vías cuando se analizan dos factores; de tres vías cuando se analizan tres factores, etc. Si se tiene un factor y una variable de agrupación —diseño de bloques— el ANOVA también es de dos vías. Si se tiene un factor y dos variables de agrupación —diseño de cuadrado latino— el ANOVA será de tres vías, y así sucesivamente.

Por otro lado, de acuerdo con el número de repeticiones se tiene ANOVA con grupos iguales para diseños balanceados; es decir, igual número de repeticiones por tratamiento, o ANOVA con grupos desiguales para diseños desbalanceados.

La forma de seleccionar los niveles de los factores también determina el tipo de análisis. Si los niveles son fijos o escogidos previamente por el investigador, se tiene el modelo I de ANOVA o modelo de efectos fijos y el análisis se hace sobre las medias. En el modelo II o modelo de efectos aleatorios, el análisis se hace para las diferentes componentes de varianza.

El objetivo central del ANOVA es la comparación de dos o más tratamientos, cada uno de los cuales representa una población, como es usual en los experimentos comparativos. Supóngase, entonces, que se tiene un factor A con k niveles o tratamientos y que cada uno de ellos está repetido r veces en el experimento. Cada unidad experimental produce una observación o dato y_{ij} que se analiza con base en un modelo estadístico llamado modelo lineal aditivo.

El análisis de un experimento está basado normalmente en un modelo matemático para las observaciones, y este modelo matemático está relacionado con y justificado por el diseño. Se puede agregar que la aleatorización es el paso que introduce los modelos probabilísticos para las observaciones. Este hecho fue observado por primera vez por Fisher, y es considerado, actualmente, una de sus mayores contribuciones al desarrollo de la ciencia.

Para el caso particular de un DCA —efectos fijos, un factor y grupos iguales— este modelo matemático tiene la siguiente estructura:

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad i = 1, 2, \ldots, r \text{ repeticiones} \qquad (4.1)$$

$$j = 1, 2, \ldots, k \text{ tratamientos,}$$

donde y_{ij} representa el dato obtenido de la repetición i-ésima en el tratamiento j-*ésimo,* así, $y_{24} = 39$ días en la tabla 4.2; μ es la media global de las poblaciones combinadas; τ_j es el efecto producido por el tratamiento j-ésimo y ε_{ij} es el error experimental aleatorio en la observación y_{ij}.

En el modelo, cada y_{ij} es una variable aleatoria lo mismo que los errores ε_{ij}, mientras que μ y τ_j son parámetros. Para hacer estimaciones y pruebas de hipótesis sobre los parámetros, es necesario establecer unas restricciones mínimas sobre el modelo:

1. Los errores ε_{ij} son variables aleatorias independientes y tienen distribución Normal $N(0,\sigma^2)$.

2. Cada tratamiento define una subpoblación con distribución Normal $N(\mu_j,\sigma^2)$; la varianza σ^2 es constante e igual para todos los tratamientos, es decir, se requieren varianzas homogéneas entre tratamientos.

3. La media μ_j, para el tratamiento j puede descomponerse como $\mu_j = \mu + \tau_j$. Para estimar en forma individual e insesgada los parámetros μ_j, se requiere, además, que $\Sigma\tau_j = 0$.

Las hipótesis del análisis de varianza de una vía y efectos fijos se refieren a la comparación de los efectos de los tratamientos y se establecen como:

H_0: $\tau_1 = \tau_2 = \ldots = \tau_k = 0$; es decir, los tratamientos no tienen ningún efecto.

H_1: al menos un $\tau_j \neq 0$; es decir, al menos un efecto es significativo.

Según la hipótesis H_0, se tiene que $\mu_j = \mu$ para todo j, esto es, las medias de los tratamientos son iguales entre sí. Por tanto, las hipótesis del análisis de varianza pueden escribirse también —como aparecen en la mayoría de los textos— en la forma:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$; o sea, no hay diferencias entre las medias de los tratamientos.

H_1: hay por lo menos dos medias diferentes entre sí.

Es prudente observar que la hipótesis H_1: todas las medias son diferentes, no es la hipótesis alternativa de H_0, puesto que esta excluye el caso de unas medias iguales y otras diferentes.

La tabla 4.3 representa el esquema general de los datos para el ANOVA de una vía.

El punto como subíndice indica que se suma sobre el subíndice que se está reemplazando. Ejemplos:

$$y_{.j} = \sum_{i=1}^{r} y_{ij}, \quad y_{i.} = \sum_{j=1}^{k} y_{ij},$$

$$y_{.j} = (1/r)\sum_{i=1}^{r} y_{ij}, \quad y_{i.} = (1/k)\sum_{j=1}^{k} y_{ij},$$

$$\bar{y}_{..} = (1/kr)\sum_{i=1}^{r}\sum_{j=1}^{k} y_{ij}, \quad y_{.j}^2 = \sum_{i=1}^{r} y_{ij}^2. \tag{4.2}$$

Tabla 4.3 Esquema de los datos para un ANOVA de una vía

Repeticiones	*Niveles del factor A*					
	T_1	T_2	…	T_j	…	T_k
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1k}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
⋮	⋮		⋮			
i	y_{i1}	y_{i2}		y_{ij}	…	
⋮	⋮		⋮			
r	y_{r1}	y_{r2}	…	y_{rj}	…	y_{rk}
Totales	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Medias	$\bar{y}_{.1}$	$\bar{y}_{.2}$	…	$\bar{y}_{.j}$	…	$\bar{y}_{.k}$

En el caso de grupos desiguales, el tratamiento j-ésimo tendrá r_j repeticiones y $\sum_j r_j = n$. Para grupos iguales, $n = kr$.

En el modelo estadístico anterior, cada unidad experimental u_{ij} produce un dato y_{ij}. La suma $\sum_i \sum_j y_{ij}^2$ se llama suma de cuadrados sin corregir, y la suma $\sum_i \sum_j \left(y_{ij} - \bar{y}_{..}\right)^2$ se llama suma de cuadrados corregida por la media. Esta última representa la variabilidad total en los datos y es la mejor medida de variación que ha resultado apropiada para el análisis de un DCA.

Por otra parte, el nombre de *análisis de varianza* tiene su origen en la partición de la suma de cuadrados corregida por la media en dos sumas de cuadrados que son mutuamente independientes: una explica la variabilidad entre grupos o tratamientos y la otra, la variabilidad del error experimental, llamada también variación dentro de grupos o error residual.

En símbolos:

$$\sum_i \sum_j \left(y_{ij} - \bar{y}_{..}\right)^2 = \sum_i \sum_j \left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 + \sum_i \sum_j \left(y_{ij} - \bar{y}_{.j}\right)^2 \quad (4.3)$$

↓		↓		↓	
Suma total de cuadrados	=	Suma de cuadrados entre tratamientos	+	Suma de cuadrados debida al error	
↓		↓		↓	
SCT	=	*SCA*	+	*SCE*	
↓		↓		↓	
$n - 1$	=	$k - 1$	+	$n - k$	*gl*

Cada suma de cuadrados tiene asociado un número de grados de libertad, en adelante *gl*, los cuales también son aditivos. El número de *gl* para estas sumas es el

número de sumandos menos uno. Al dividir una suma de cuadrados por sus *gl*, se obtiene una varianza. En el ANOVA, estas varianzas se llaman medias cuadráticas o también cuadrados medios. En particular:

$$\frac{SCA}{k-1} = MCA,$$

es la media cuadrática de tratamientos o del factor A. De ahí el uso de la *A;* más adelante se usarán las letras B, C, etc., para representar otros factores. Por otro lado,

$$\frac{SCE}{n-k} = MCE, \tag{4.4}$$

es la media cuadrática del error o varianza residual.

Un teorema sobre sumas cuadráticas debido a W. G. Cochran, permite concluir que las medias cuadráticas MCA y MCE que tienen distribución *ji*-cuadrado, son probabilísticamente independientes, por tanto, el cociente $F = MCA / MCE$ tiene una distribución $F_{\nu 1, \nu 2}$ con $\upsilon_1 = k - 1$ y $\upsilon_2 = n - k$ *gl* bajo la hipótesis nula. Esta conclusión es la base para establecer la regla de decisión de las hipótesis del ANOVA, la cual se enuncia como:

Rechazar H_0 si $F > F_{\alpha, \upsilon 1, \upsilon 2}$.

Se verá ahora por qué al comparar varianzas es posible concluir sobre la igualdad o la diferencia de medias. Por métodos probabilísticos se encuentra que el valor esperado de la MCE es σ^2, o sea que la MCE es un estimador insesgado de la varianza del error. También se encuentra que el valor esperado de la MCA es $\sigma^2 + r\,\Sigma\tau_j^2 / (k-1)$.

Si el efecto τ_j es cero, entonces $\Sigma\tau_j^2 = 0$ y la *MCA* también estima a σ^2. Bajo la hipótesis nula, el cociente F es igual a 1.0, pero si algún τ_j es diferente de cero, entonces $F > 1.0$, ya que el numerador será mayor que el denominador. Un valor de F cercano a la unidad indica que los τ_j son cero o muy pequeños, lo que implica que $\mu_j = \mu$, o sea, las medias de los tratamientos pueden suponerse iguales entre sí y, por tanto, se aceptaría H_0. Al contrario, un valor grande de F indica que algunos τ_j, son diferentes de cero y que las correspondientes medias μ_j son diferentes entre sí, por lo que deberá rechazarse H_0.

Toda la información sobre el ANOVA se recoge y se resume, usualmente, en una tabla que tiene una estructura como la de la tabla 4.4.

Otra discusión interesante sobre los principios del ANOVA, que se recomienda al lector, se encuentra en el libro de Schefler (1981). Un enfoque diferente, donde se enfatiza el tratamiento gráfico del ANOVA, se encuentra en Box, Hunter y Hunter (2005).

4.3 Inferencia en ANOVA

Con los resultados del ANOVA se pueden hacer estimaciones y probar hipótesis a

Tabla 4.4 Análisis de varianza de una vía y efectos fijos

Fuente de variación	*Suma de cuadrados*	*Grados de libertad*	*Medias cuadráticas*	*Valor esperado de la MC*	*Cociente F*	*Probabilidad p*
Entre tratamientos (factor A)	SCA	$k-1$	MCA	$\sigma^2 + \frac{r}{k-1}\sum \tau_j^2$	$\frac{MCA}{MCE}$	0.xx
Dentro de tratamientos (error residual)	SCE	$n-k$	MCE	σ^2		
Total	SCT	$n-1$				

posteriori. En primer lugar, la prueba F permite decidir si hay o no hay diferencias entre medias de tratamientos. Este es un paso intermedio en el análisis, ya que al concluir que existen diferencias, el paso siguiente es encontrar las medias que son significativamente diferentes entre sí y estimar su diferencia.

Para hallar el cociente F correspondiente a los datos de la tabla 4.2, se calculan las sumas de cuadrados con las fórmulas siguientes, y usando una calculadora. Antes de hacer estos cálculos es conveniente constatar que se cumplan los supuestos del ANOVA. Como este es el tema del capítulo siguiente, no se hace el chequeo todavía, pero se continúa con el análisis aceptando que los supuestos se cumplen.

Término de corrección; $C = (1/n)\left(\sum_i \sum_j y_{ij}\right)^2 = (1/20)(899)^2 = 40410.$

$$SCT = \sum_i \sum_j y_{ij}^2 - C.$$

$$SCT = (41)^2 + (44)^2 + \ldots + (46)^2 + (41)^2 - C = 40661 - 40410 = 251.$$

$$SCA = (1/r)\sum_j \left(\sum_i y_{ij}\right)^2 - C.$$

$$SCA = (1/5)(215^2 + 240^2 + 232^2 + 212)^2 - 40410 = 108.6.$$

$$SCE = \sum_i \sum_j y_{ij}^2 - (1/r)\sum_j \left(\sum_i y_{ij}\right)^2.$$

$$SCE = 40661 - 40518.6 = 142.4.$$

Considerando la aditividad de las sumas de cuadrados, resulta más corto calcular la última suma como la diferencia:

$$SCE = SCT - SCA = 251 - 108.6 = 142.4.$$

Entonces:

$$F = \frac{MCA}{MCE} = \frac{108.6/3}{142.4/16} = 4.07.$$

Al comparar este resultado con el valor tabulado $F_{0.05,3,16} = 3.24$, se rechaza H_0 y se concluye que el promedio de días de permanencia de los pacientes en el hospital es diferente dependiendo del tratamiento empleado. El nivel de significación de la prueba es $p = 0.0252$ dado por el computador en una tabla similar a la 4.5.

Aunque teóricamente se espera que $F \geq 1.0$, en la práctica pueden darse valores de $F < 1.0$, lo cual indica que no se cumplieron ciertos supuestos en el proceso experimental (Ostle, 1988).

El ANOVA anterior permite concluir que hay diferencias estadísticamente significativas ($p < 0.05$) entre los tratamientos. Sin embargo, la conclusión desde el punto de vista clínico podría ser que las diferencias no son relevantes, pero esta afirmación no contradice el resultado estadístico puesto que se origina en otro contexto, o sea, en el campo de la medicina.

La comparación de todas las medias para establecer las diferencias entre ellas exige consideraciones de diseño que se verán en el capítulo 6.

Tabla 4.5 Tabla del análisis de varianza para los datos del ejemplo 4.1

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Terapias	108.6	3	36.2	4.07	0.025
Error	142.4	16	8.9		
Total	251.0	19			

Conocido el modelo estadístico es posible hacer estimaciones de sus parámetros. Las siguientes estimaciones son generales; algunas no se aplican a todos los problemas, pero se presentan como una guía para el lector.

1. Una estimación de la varianza del error experimental es $\hat{\sigma}^2 = MCE = 8.9$. Esta es una varianza mancomunada del experimento; es muy útil para el análisis siempre que las varianzas de los tratamientos sean homogéneas.

2. Una estimación puntual de la media global del experimento es $\hat{\mu} = \bar{y}_{..} = 899/20 = 45.0$ días. Tiene sentido estimar esta media solo cuando se acepta H_0.

3. Las medias de los tratamientos se estiman a partir de las medias muestrales o medias de grupo. Para el tratamiento j-ésimo la estimación es $\hat{\mu}_j = \bar{y}_{.j}$. Entonces,

$$\hat{\mu}_1 = \bar{y}_{.1} = 215/5 = 43.0 \text{ días},$$

$$\hat{\mu}_2 = \bar{y}_{.2} = 240/5 = 48.0 \text{ días},$$

$$\hat{\mu}_3 = \bar{y}_{.3} = 232/5 = 46.4 \text{ días},$$

$$\hat{\mu}_4 = \bar{y}_{.4} = 212/5 = 42.4 \text{ días}.$$

4. Los efectos de los tratamientos del factor A son las diferencias $\tau_j = \mu_j - \mu$, las cuales se estiman mediante $\hat{\tau}_j = \hat{\mu}_j - \hat{\mu}$. Individualmente, sus estimaciones son:

$$\hat{\tau}_1 = \hat{\mu}_1 - \hat{\mu} = 43.0 - 45.0 = -2.0 \text{ días},$$

$\hat{\tau}_2 = \hat{\mu}_2 - \hat{\mu} = 48.0 - 45.0 = +3.0$ días,

$\hat{\tau}_3 = \hat{\mu}_3 - \hat{\mu} = 46.4 - 45.0 = +1.4$ días,

$\hat{\tau}_4 = \hat{\mu}_4 - \hat{\mu} = 42.4 - 45.0 = -2.6$ días,

Por ejemplo, el resultado −2.0 días para el efecto del primer tratamiento, significa que los pacientes tratados con la primera terapia permanecen en el hospital dos días menos que el promedio; es decir, el efecto del primer tratamiento es disminuir la permanencia en el hospital en un promedio de dos días.

El quinto paciente del tratamiento T_1 permaneció 42 días hospitalizado. Este dato, y en forma similar para los demás, puede descomponerse según el modelo lineal, como:

$y_{51} = \mu + \tau_1 + \varepsilon_{51}$, esto es:

$42 = 45.0 - 2.0 - 1.0$, por tanto, $\hat{\varepsilon}_{51} = -1.0$.

5. El residuo negativo (−1.0 días) es un estimativo del error aleatorio ε_{51} en el paciente antes citado. En otras palabras, si no hubiese error aleatorio, los pacientes con tratamiento T_1 habrían permanecido $45.0 - 2.0 = 43.0$ días hospitalizados. Este paciente en particular, permaneció 42 días; el hecho de haber salido un día antes no es explicable por el tratamiento recibido sino por otras causas no identificadas por el investigador.

6. Un intervalo de confianza para la media μ_j se calcula con la fórmula $\bar{y}_{.j} \pm t_{\alpha/2,\nu}\sqrt{MCE/r}$, donde ν son los *gl* asociados con la *MCE*, esto es, $\nu = n - k$. Por ejemplo, un intervalo del 95% de confianza para la media del tratamiento T_4 es:

$\bar{y}_{.4} \pm t_{0.05,16}\sqrt{MCE/r}$.

$42.4 \pm 2.12\sqrt{8.9/5}$.

$39.6 < \mu_4 < 45.2$.

7. Un intervalo de confianza para la diferencia de dos medias $\mu_i - \mu_j$ tiene la forma $\left(\bar{y}_{.i} - \bar{y}_{.j}\right) \pm t_{\alpha/2,\nu}\sqrt{2\,MCE/r}$.

El intervalo del 90% de confianza para $\mu_2 - \mu_1$, es:

$(48.0 - 43.0) \pm t_{0.05,16}\sqrt{2 \times 8.9/5}$.

$5.0 \pm 1.746 \times 1.887$.

$1.7 < \mu_2 - \mu_1 < 8.3$.

El tratamiento T_2 es menos efectivo que el tratamiento T_1 —mayor permanencia en el hospital— y la diferencia está entre 1.7 y 8.3 días. Este intervalo no contiene el cero, lo que permite concluir que los tratamientos T_1 y T_2 tienen promedios diferentes.

Es de anotar que este procedimiento no debería seguirse para comparar todas las medias por pares. Las seis comparaciones serían conjuntamente dependientes y, por tanto, los resultados no serían válidos estadísticamente. El procedimiento de comparaciones por pares merece un análisis más detallado, y por eso se ha pospuesto para el capítulo sexto, el cual trata sobre comparaciones múltiples.

4.4 ANOVA para grupos desiguales

El diseño completamente aleatorizado no tiene restricciones sobre el número de repeticiones por tratamiento. No obstante, se recomienda diseñar igual número de repeticiones con el objeto de aumentar la potencia de las pruebas. Ahora bien, cuando se tienen grupos iguales se dice que el diseño está balanceado. En algunos experimentos este balanceo se destruye a causa de, por ejemplo, la pérdida de unidades experimentales o la imposibilidad de consecución de suficientes unidades experimentales homogéneas. El ANOVA para el diseño desbalanceado es similar al de grupos iguales; solo varía en algunas fórmulas de cálculo.

Ejemplo 4.2 Para comparar cuatro dietas D_1, D_2, D_3 y D_4, respecto a su influencia en el tiempo de coagulación de la sangre, se seleccionaron 24 animales y cada uno recibió aleatoriamente una de las dietas. Los resultados se dan en la tabla 4.6 (adaptado de Box, Hunter y Hunter, 2005).

Tabla 4.6 Tiempo en segundos de coagulación de la sangre

	Dietas				
	D_1	D_2	D_3	D_4	
	62	63	68	56	
	60	67	66	62	
	63	71	71	60	
	59	64	67	63	
		65	68	63	
		66	68	64	
				63	
				59	
Totales	244	396	408	488	1536
Medias	61	66	68	61	64
r_j	4	6	6	8	24

Los cálculos para construir la tabla 4.7 del ANOVA, son:

$$n = \sum_j r_j = 24.$$

$$C = (1/n)\left(\sum_i^k \sum_j^{r_j} y_{ij}\right)^2 = (1\,536)^2/24 = 98\,304.$$

$$SCT = \sum_i \sum_j y_{ij}^2 - C = 98644 - 98304 = 340.$$

$$SCA = \left(\sum_j y_{.j}^2 \Big/ r \right) - C = \frac{(244)^2}{4} + \frac{(396)^2}{6} + \frac{(408)^2}{6} + \frac{(488)^2}{8} - C$$

$$= 98532 - 98304 = 228.$$

Estos resultados permiten concluir que hay diferencias significativas entre dietas, pero se dejará hasta el ejemplo 6.4 sobre comparaciones múltiples, el análisis de las comparaciones entre ellas.

Algunas estimaciones de parámetros son las siguientes:

$$\bar{y}_{.j} \pm t_{\alpha/2,\upsilon} \sqrt{MCE / r_j}\,.$$

1. El intervalo de confianza para una media es:

Aplicado a μ_4, por ejemplo, y con 95% de confianza, se tiene:

$$\bar{y}_{.4} \pm t_{0.025,20} \sqrt{5.6/8} = 61 \pm 2.086 \times 0.8367 = 61 \pm 1.7 \text{ segundos.}$$

2. Las medias de las dietas D_3 y D_4 son diferentes, como puede deducirse del siguiente intervalo del 95% de confianza:

$$(\bar{y}_{.3} - \bar{y}_{.4}) \pm t_{\alpha/2,\nu} \sqrt{\frac{MCE}{r_3} + \frac{MCE}{r_4}}\,.$$

Dando valores numéricos se tiene:

$$(68 - 61) \pm 2.086\sqrt{(5.6/6) + (5.6/8)} = 7 \pm 2.7, \text{ o también,}$$

$$4.3 < \mu 3 - \mu 4 < 9.7.$$

4.5 Modelo de componentes de varianza

La selección de los tratamientos o niveles del factor de estudio determina el procedimiento que debe usarse en el análisis. El modelo I es apropiado cuando los tratamientos son fijos o el experimentador los elige por su importancia, disponibilidad o comodidad en el manejo. Por ejemplo, tratamientos médicos, variedades, concentraciones, lugares y tiempos. En este modelo, los parámetros son generalmente las medias de los tratamientos, y el objetivo es la comparación y estimación de estos promedios.

En el modelo II, los niveles del factor de estudio son una población de la cual se toma una muestra al azar; por ejemplo, animales de una especie dada, hojas de una planta, sitios de un lago o personas con una característica definida. Cada individuo

Tabla 4.7 Tabla del ANOVA para los datos del ejemplo 4.2

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Dietas	228	3	76.0	13.6	<0.001
Error	112	20	5.6		
Total	340	23			

seleccionado en la muestra origina una o varias lecturas de la variable de respuesta, sea en días diferentes u otra situación que permita observar la variabilidad de la característica en el mismo individuo. Los parámetros de interés en este modelo son las varianzas, y el objetivo primordial es estimar sus magnitudes absolutas y relativas.

El modelo estadístico para un DCA —un factor y efectos aleatorios— tiene la forma:

$$y_{ij} = \mu + A_j + \varepsilon_{ij}, \text{ con } i = 1, 2, \ldots, r_j \text{ y } j = 1, 2, \ldots, k. \quad (4.5)$$

La diferencia con respecto al modelo de efectos fijos está en el segundo término. Aquí los efectos A_j son variables aleatorias, mientras que en el modelo I los efectos τ_j son parámetros. Para analizar este diseño, deben introducirse las siguientes restricciones:

1. Los efectos A_j son variables aleatorias independientes con distribución Normal $N(0,\sigma_A^2)$, donde σ_A^2 es la varianza de los valores medios de la variable de respuesta para todos los posibles niveles del factor.
2. Los errores ε_{ij} son variables aleatorias independientes con distribución Normal $N(0,\sigma^2)$.
3. Las variables A_j y ε_{ij} son variables aleatorias independientes entre sí.

La tabla del ANOVA (véase tabla 4.8) es similar a la del modelo I, pero se diferencian en las medias cuadráticas esperadas. En el modelo de componentes de varianza la *MCA* estima una combinación lineal de las dos varianzas, es decir, estima la varianza del error (σ^2) más la varianza entre las medias de todos los grupos posibles del factor aleatorio (σ_A^2). Para analizar si el factor de estudio tiene algún efecto en la variable de respuesta, se prueban las hipótesis siguientes:

H_0: $\sigma_A^2 = 0$, la varianza entre las medias de todos los grupos del factor A es cero.

H_1: $\sigma_A^2 > 0$.

Si $F > F_{\alpha, v1, v2}$, se rechaza H_0. La regla de decisión es la misma que la del modelo I, pero la interpretación es bastante diferente. Los datos del modelo II provienen de un diseño distinto, el objetivo es estimar varianzas y no medias; además, el espacio inferencial es mucho más amplio para este modelo. Entonces, aunque el ANOVA sea el mismo, las conclusiones son diferentes puesto que los diseños experimentales son diferentes, según Lindman (1992).

Tabla 4.8 Tabla del ANOVA de un factor y efectos aleatorios

Fuente de variación	*SC*	*gl*	*MC*	*MC esperada*	*F*	*p*
Factor A	SCA	$k-1$	MCA	$\sigma^2 + r_0\sigma_A^2$	$\frac{MCA}{MCE}$	0.xx
Error	SCE	$n-k$	MCE	σ^2		
Total	SCT	$n-1$				

$r_0 = \left(n - \sum_i r_i^2/n\right) \Big/ (k-1), \; n = \sum_j r_j$

Tabla 4.9 Aumento diario de peso en kg según el ejemplo 4.3

Toros			
1	*2*	*3*	*4*
1.46	1.17	0.98	0.95
1.23	1.08	1.06	1.10
1.12	1.20	1.15	1.07
1.23	1.08	1.11	1.11
1.02	1.01	0.83	0.89
1.15	0.86	0.86	1.12
	1.19	0.99	1.15
	0.97		1.10

Ejemplo 4.3 La tabla 4.9 contiene los datos de la ganancia promedio diaria de peso en terneros, cuyos padres fueron cuatro toros de la misma raza. Los terneros fueron alimentados de la misma manera y durante el mismo periodo de tiempo. Se desea saber si la variabilidad de la ganancia en peso es atribuible a la diferencia entre los padres o si es propia de los terneros, los cuales son tratados como unidades experimentales (adaptado de Gill, 1978).

Suponiendo que los cuatro toros fueron seleccionados al azar de un número grande de posibles padres, los datos se analizan mediante un ANOVA modelo II. Las hipótesis son:

H_0: $\sigma_A^2 = 0$; no hay variabilidad entre padres.

H_1:$\sigma_A^2 > 0$; la variabilidad entre padres es significativa.

En esta investigación no se consideran los promedios por padre —cada toro es un tratamiento— puesto que los cuatro toros del experimento no son los únicos que interesan; el objetivo se centra en la variabilidad poblacional de los padres y no en la comparación de los cuatro de la muestra, que sería el objetivo en el modelo I. El factor de estudio es la paternidad o influencia hereditaria, que se expresa en el aumento de peso de los terneros hijos. Aquí, cada ternero es una unidad experimental, pero en otros estudios cada animal podría generar varias unidades experimentales correspondientes a lecturas o mediciones en tiempos diferentes.
La tabla del ANOVA se construye de la misma forma que para el modelo I y los resultados se dan en la tabla 4.10.

El valor tabulado $F_{0.05,3,25} = 2.99$ es menor que el cociente F, lo que permite concluir que la variabilidad entre padres es significativa estadísticamente. A partir de este resultado, es interesante conocer qué parte de la variabilidad en las unidades experimentales es asignable a los padres, σ_A^2, y qué parte es atribuible al error experimental, σ^2. Es obvio que no tendría sentido diferenciar estas variabilidades si la conclusión del ANOVA hubiese sido H_0.

Si se satisfacen los supuestos del modelo II, como se verá en el capítulo 5, la varianza de cada observación y_{ij} puede expresarse mediante dos componentes dados por los términos de la suma $\sigma_y^2 = \sigma_A^2 + \sigma^2$. En el modelo se tienen tres parámetros ($\mu.$,σ_A^2, σ^2) y sus estimaciones son las siguientes:

1. Las estimaciones de las componentes de varianza se obtienen a partir de las medias cuadráticas, así:

$$\hat{\sigma}^2 = MCE = 0.0142.$$

$$r_0 = [29 - (6^2 + 8^2 + 7^2 + 8^2)/29]/3 = 7.22.$$

$$\hat{\sigma}^2 + r_0\hat{\sigma}_A^2 = MCA, \text{ luego } \hat{\sigma}_A^2 = (MCA - MCE)/r_0.$$

$$\hat{\sigma}_A^2 = (0.0468 - 0.0142)/7.22 = 0.0045, \text{ entonces}$$

$$\hat{\sigma}_y^2 = 0.0045 + 0.0142 = 0.0187.$$

Normalmente, estas estimaciones no dan una información fácilmente interpretable; por esto, lo usual es expresar los resultados como porcentajes de la variación total. El cociente $r_I = \hat{\sigma}_A^2/\hat{\sigma}_y^2 = 0.0045/0.0187 = 0.2406$ mide la proporción de la variabilidad total que es atribuible a la variabilidad del factor; en este caso, a la variabilidad entre

padres. En porcentaje, esta variabilidad es el 24.1%; el otro 75.9% es la variabilidad entre terneros del mismo padre.

2. El cociente r_1 se conoce como coeficiente de correlación intraclase y es una medida de la semejanza de los individuos en un mismo grupo. Este coeficiente varía entre 0 y 1; entre más grande sea su valor, mayor es la similitud entre los elementos de cada grupo. La significancia de la correlación intraclase está ligada directamente a la prueba F del ANOVA. En efecto, si se rechaza H_0 se concluye que la variabilidad causada por los padres es alta y que la variabilidad restante, causada por otros factores como el ambiente, es baja, lo que lleva a la similitud entre las observaciones por grupo.

Tiene sentido hablar de la correlación intraclase porque en el modelo II, a diferencia del modelo I, las observaciones y_{ij} en un mismo grupo pueden ser estadísticamente dependientes (Scheffé, 1999).

3. El coeficiente de heredabilidad h^2, de gran importancia para los genetistas, se estima como $\hat{h}^2 = 4r_1$, este es una medida del grado de influencia genética que puede recibir un genotipo. Para el caso de la ganancia diaria de peso en terneros, este coeficiente varía entre 0.3 y 0.5, según Gardner y otros (1991). Para el ejemplo, la estimación del coeficiente de heredabilidad dio 0.96, un valor demasiado alto pero explicable por la gran variabilidad de $\hat{\sigma}_A^2$.

4. La distribución de $\hat{\sigma}_A^2$ es sesgada y muy variable, por tanto, los intervalos de confianza para σ_A^2 son aproximados. Estos se han derivado para el caso de grupos iguales, pero pueden extenderse, agregando otra aproximación, a grupos desiguales con r_0 en vez de r. Uno de los intervalos, propuesto por S. Moriguti en 1954 y citado por Gill (1978, vol. 1), tiene los siguientes límites:

$$\text{Límite inferior} = \frac{MCE}{r_0}\left[\frac{F}{F_\infty} - 1 + \frac{F_\upsilon}{F}\left(1 - \frac{F_\upsilon}{F_\infty}\right)\right].$$

$$\text{Límite superior} = \frac{MCE}{r_0}\left[F F'_\infty - 1 + \frac{1}{F F'_\upsilon}\left(1 - \frac{F'_\infty}{F'_\upsilon}\right)\right].$$

Donde $F_\upsilon = F_{\alpha/2,\upsilon1,\upsilon2}$, $F'_\upsilon = F_{\alpha/2,\upsilon2,\upsilon1}$, $F_v = (MCA / MCE)$, $F_\infty = F_{\alpha/2,\upsilon1,\infty}$, $F'_\infty = F_{\alpha/2,\infty,\upsilon1}$, con $\upsilon_1 = k - 1$, $\upsilon_2 = n - k$.

Un intervalo de confianza del 95% para σ_A^2 tendrá:

El límite inferior resultó negativo, pero como el parámetro estimado por este resultado no puede ser negativo, entonces se redefine el límite dándole el valor de cero. El intervalo será $0 \leq \sigma_A^2 \leq 0.0882$, que resulta demasiado amplio si se observa que el límite superior es casi 20 veces la estimación puntual de 0.0045 obtenida anteriormente. Una de las razones que explican la poca precisión en la estimación de σ_A^2 es el reducido tamaño de la muestra analizada.

Algunas estimaciones de las componentes de varianza pueden resultar negativas —cuando $MCA < MCE$—; las principales causas de este resultado son la variación muestral de las medias cuadráticas, la falta de aleatorización entre tratamientos y la correlación

Tabla 4.10 Tabla del ANOVA para el ejemplo 4.3

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Entre toros	0.1403	3	0.0468	3.29	0.037
Error	0.3555	25	0.0142		
Total	0.4958	28			

negativa de los errores experimentales dentro de algún grupo, Gill (1978, vol. 1). Las estimaciones negativas de una varianza no tienen interpretación real. Una alternativa es reemplazar el valor negativo por cero; sin embargo, es más prudente revisar y corregir el diseño utilizado.

5. Un intervalo de confianza para el coeficiente de correlación intraclase poblacional ρ_I puede calcularse como:

$$\frac{I}{I+1} \leq \rho_I \leq \frac{S}{S+1},$$

$$\text{donde } I = \frac{1}{k}\left(\frac{F}{F_{\alpha/2,\upsilon_1,\upsilon_2}} - 1\right),\ S = \frac{1}{k}\left(\frac{F}{F_{1-\alpha/2,\upsilon_1,\upsilon_2}} - 1\right).$$

Para $\alpha = 0.05$, $\upsilon_1 = 3$, $\upsilon_2 = 25$, se tiene $F_{0.025,3,25} = 3.69$ y $F_{0.975,3,25} = 0.071$. Luego $I = [(3.29 / 3.69) - 1] / 4 = -0.0264$ y $S = [(3.29 / 0.071) - 1] / 4 = 11.33$. Entonces, el intervalo es $-0.027 \leq \rho_I \leq 0.919$. Aunque el coeficiente poblacional ρ_I es siempre positivo, se prefiere no modificar el intervalo hallado para no dar la impresión de un intervalo más corto, Scheffé (1999). Se observa, también, la imprecisión de este intervalo, explicable por su dependencia de las componentes de varianza muestrales.

6. Otra diferencia del modelo de componentes de varianza con respecto al modelo de efectos fijos, se da en el valor esperado μ, el cual es igual para todas las observaciones. Esta media, μ, se estima puntualmente mediante $\bar{y}_{..} = 31.24/29 = 1.08$ kg; y por intervalos con la fórmula $\bar{y}_{..} \pm t_{\alpha/2,\nu}\sqrt{MCA/n}$, con $\upsilon = k - 1$. Para una confiabilidad del 95%, este intervalo será:

$$1.08 \pm 3.18\sqrt{0.0468/29}, \text{ es decir, } 1.08 \pm 0.13 \text{ kg.}$$

4.6 Potencia y tamaño muestral en ANOVA

La técnica principal ofrecida por la teoría estadística para decidir sobre el número de repeticiones necesarias en un experimento, es el cálculo de la potencia de las pruebas que se van a realizar. Para la prueba *F* del ANOVA de una vía, el cálculo directo de la potencia es generalmente complejo, pero se han construido algunas gráficas, llamadas curvas características de operación, que permiten estimar un valor para la probabilidad β o error de tipo II. La potencia $1 - \beta$ se deduce a partir de esta probabilidad.

Número de repeticiones en el modelo I

Las gráficas más comunes, construidas por Pearson y Hartley, se encuentran reproducidas en varios textos aunque con ligeras modificaciones, por ejemplo, en Neter y otros (1996), Scheffé (1999) y Montgomery (2007). Se utilizarán las que trae este último autor porque son más completas.

Las curvas fueron construidas para dar el valor de β en la ordenada cuando se proponen valores de un parámetro Φ sobre la abscisa y se suponen valores conocidos de α, υ_1 y υ_2. El parámetro Φ se llama parámetro de no centralidad de la distribución *F*, y es una medida del grado de desigualdad de las μ_j, a mayor diferencia entre ellas más alto es el valor de Φ. Una estimación de Φ se obtiene mediante la expresión:

$$\Phi = \sqrt{\frac{r\sum_j \tau_j^2}{k\sigma^2}}.$$

Hay gráficas disponibles solamente para valores de $\alpha = 0.05$ y 0.01, y *gl*: $\upsilon_1 = 1, 2, \ldots, 8$ y $\upsilon_2 = 6, 7, 8, 9, 10, 12, 15, 20, 30, 60, \infty$; para otros valores de los *gl* es suficiente una interpolación visual. Para $\upsilon_1 > 8$ véase Gill (1978, vol. 1).

Ejemplo 4.4 Calcular el número de repeticiones para un experimento similar al del ejemplo 4.1, suponiendo que la potencia no debe ser inferior a 0.80.

Se comienza dando los mismos valores del ejemplo citado a los parámetros, así:

$$\alpha = 0.05, \upsilon_1 = 3, \upsilon_2 = 16, \sigma^2 = 8.9, \Sigma\tau_j^2 = 4 + 9 + 1.96 + 6.76 = 21.72.$$

Entonces, $\phi^2 = (5 \times 21.72) / (4 \times 8.9) = 3.05$ y $\phi = 1.75$.

En el apéndice B y en la gráfica de $\nu_1 = 3$ y $\alpha = 0.05$, se localiza $\phi = 1.75$, desde este punto se sube hasta cortar la línea $\nu_2 = 15$ y luego, se va horizontalmente hacia la izquierda y se lee la probabilidad $\beta \approx 0.28$. La potencia para el ejemplo 4.1 es, por tanto, de 0.72.

Si se desea realizar un experimento como el anterior pero con una potencia mayor, se debe aumentar el número de repeticiones *r*. Calcular nuevamente la potencia con $r = 6$. Para este valor, $\phi = 1.91$ y $\beta \cong 0.19$; además la potencia sería de 0.81, que cumple con la condición inicial.

En el ejemplo anterior, el procedimiento resulta inmediato porque se conocen todos los parámetros. En una situación real, el problema es más complejo ya que ϕ no es calculable directamente. El valor de *r* es precisamente el valor que se desea conocer; la varianza σ^2 es igualmente desconocida, y los τ_j no son estimables antes de realizar el experimento.

El primer paso es la estimación de σ^2. Si se conocen experimentos similares, se puede reemplazar σ^2 por *MCE*. En otro caso, debe hacerse una estimación razonable mediante la información que posea el investigador, pues se supone que este no parte de cero, sino que tiene algún conocimiento previo del comportamiento del fenómeno.

El segundo es la estimación de τ_j. Si fuera posible proponer valores para los τ_j, se calcularía la suma de los cuadrados como en el ejemplo anterior. Lo más común es no tener valores individuales para estos efectos, entonces, otra alternativa es que el investigador especifique la diferencia mínima que considera significativa entre las dos medias poblacionales más distantes y, con base en este valor, llamado Δ, calcular un valor tentativo para ϕ mediante la siguiente fórmula, Scheffé (1999):

$$\phi = \sqrt{\frac{r\Delta^2}{2k\sigma^2}}, \text{ en general } \sum \tau_j^2 \geq \frac{\Delta^2}{2}.$$

El tercer y último paso es la estimación de *r*. Para esta estimación, hay que proceder un poco a la inversa. Se parte de un valor *r* más o menos grande y se encuentra el valor de ϕ correspondiente a una potencia establecida. Conocido el valor de ϕ, se despeja *r* de la fórmula anterior y se obtiene una primera aproximación a su valor. Luego se estima la potencia correspondiente a este nuevo valor de *r*; si resulta menor que la establecida, debe aumentarse dicho valor, y si resulta mayor, debe disminuirse.

Considérese el siguiente caso: el estudio detectará como significativa una diferencia entre medias igual o mayor que cuatro días, es decir, $\Delta = 4$ días. Supóngase que se estimó la varianza en $\hat{\sigma}^2 = 8.9$; hay $k = 4$ tratamientos para comparar y se requiere una potencia de 0.80.

Entonces $\phi^2 = 16\,r / 8 \times 8.9 = 0.225r$.

Se inicia con $r = 15$. En la gráfica de $\nu_1 = 3$ y $\alpha = 0.05$, se localiza $\beta = 0.20$ y para

$\nu_2 = 4 \times 14 = 56$ se lee $\phi \approx 1.72$. Entonces, $\phi^2 = 2.96$ y el r despejado es 13.15. Se repite el proceso con $r = 13$, de modo que $\phi = \sqrt{0.225 \times 13} = 1.71$ y $\nu_2 = 4 \times 12 = 48$. Estos valores dan una probabilidad $\beta \approx 0.20$. Se concluye que se requieren 13 pacientes como mínimo por tratamiento, si se desea una potencia de 0.80 para la prueba F, suponiendo que diferencias entre medias poblacionales de cuatro días o mayores son significativas.

El número de repeticiones r es muy sensible a la variación en los parámetros Δ y σ, como también en la potencia requerida. Dejando fijos a α y ν_1, se puede ver, en la tabla 4.11, la variación de r.

Al aumentar el número de repeticiones r, también aumenta la potencia. Para una potencia fija, r se puede disminuir si ϕ se aumenta. Pero el parámetro ϕ depende básicamente del cociente Δ/σ, el cual puede aumentar, sea porque la varianza es pequeña o porque la diferencia significativa se supone grande. Por otra parte, diferencias grandes entre las μ_j son fáciles de detectar con pocos datos. Como no hay mucha precisión en cuanto a la varianza estimada, es aconsejable investigar varios tamaños muestrales, entre un rango probable de valores de σ^2, antes de decidirse por un tamaño definitivo.

Otros enfoques interesantes sobre la estimación del tamaño muestral en ANOVA, se encuentran, por ejemplo, en Sokal y Rohlf (1995) y en Quinn y Keough (2003).

Número de repeticiones en el modelo II

La potencia de la prueba F del ANOVA para el modelo de efectos aleatorios se basa en la distribución F, pero la determinación de su valor resulta más simple usando las curvas características de operación. Estas curvas, como las del modelo I, dan valores de β en la ordenada para diferentes valores del parámetro λ en la abscisa.

$$\lambda = \sqrt{1 + \frac{r\sigma_A^2}{\sigma^2}}. \tag{4.5}$$

Ejemplo 4.5 Calcular la potencia de la prueba F para los datos del ejemplo 4.3.

Los parámetros conocidos de este ejemplo son: $\alpha = 0.05$, $\upsilon_1 = 3$, $\upsilon_2 = 25$, $\sigma_A^2 = 0.0045$, $\sigma^2 = 0.0142$ y $r_0 = 7.22$. El método es válido únicamente para grupos iguales, pero se hace la aproximación con r_0 para conservar los valores del ejemplo citado. Con estos datos, el parámetro $\lambda = 1.81$. En la gráfica del apéndice C, se lee, para este λ, una probabilidad $\beta \approx 0.57$, o sea que la potencia es 0.43. La potencia calculada es muy baja a causa del reducido número de repeticiones y, además, por tener grupos desiguales. Si se quiere repetir este experimento fijando una potencia de 0.80 y con los mismos parámetros, el cálculo de r para grupos iguales sería el siguiente.

Se fija un valor para r, por ejemplo 10, entonces $\upsilon_2 = 4 \times 9 = 36$. Con $\beta = 0.20$, se lee en la gráfica el valor $\lambda \approx 3.1$. Se despeja r de la fórmula 4.5 y se obtiene:

$$r = \frac{\left(\lambda^2 - 1\right)\sigma^2}{\sigma_A^2} = 27.17.$$

Se ensaya con $r = 27$ y se calcula un nuevo $\lambda = 3.09$. Con $\upsilon_2 = 104$, se lee $\beta \approx 0.18$. Como β puede ser hasta de 0.20, se disminuye r a 24 y se hace otro ensayo. Para este r se tiene $\lambda = 2.93$, $\nu_2 = 92$ y $\beta \approx 0.20$. Así que para lograr una potencia de 0.80, se requiere mínimo 24 repeticiones por tratamiento.

Los tamaños de muestra obtenidos son aproximados, ellos dependen tanto de las estimaciones de los parámetros como de las lecturas visuales en las gráficas, las cuales pueden variar de una persona a otra.

Tabla 4.11 Valores de r para diferentes valores de los parámetros Δ, σ y $1 - \beta$

Δ	σ^2	$1-\beta$	r	ϕ	β
4	8.9	0.80	13	1.71	0.20
3	8.9	0.80	22	1.66	0.20
4	10.0	0.90	15	1.72	0.20
4	8.9	0.80	17	1.96	0.09
5	6.0	0.80	7	1.91	0.18

Las siguientes observaciones sobre el número de repeticiones complementan la discusión anterior y sirven de guía para estudios posteriores.

1. En condiciones similares, el tamaño muestral es mayor para el modelo II que para el modelo I, porque en el modelo II se estiman parámetros cuadráticos mientras que en el modelo I los parámetros son lineales.

2. En lo posible, deberán diseñarse experimentos con igual número de repeticiones por grupo, sobre todo si se planean comparaciones de medias por pares; así se obtiene mayor precisión en las comparaciones y mayor potencia.

3. Conocido el tamaño del efecto, este se detecta más fácilmente con una muestra grande.

4. Si el número de repeticiones se ha fijado previamente, el cálculo de la potencia para el ANOVA permite ahorrar tiempo y esfuerzo, ya que para una potencia baja es preferible no realizar el experimento, pues la probabilidad de detectar un efecto importante, si este existe, es mínima.

La situación anterior tiene un paralelo con el caso de la persona que desea pesar, por ejemplo, una moneda y solo dispone de una balanza graduada de 100 en 100 g. Puede efectuar la medición pero el resultado es tan incierto que sería casi igual a no usar dicha balanza.

5. Si el experimento ya ha sido realizado, es útil preguntarse por la potencia para un tamaño de efecto prefijado, pero calcular la potencia para el efecto observado carece de interés.

6. Puede ocurrir que el número total de unidades experimentales esté limitado y que no todos los tratamientos puedan incluirse en el experimento. El problema se convierte en determinar el valor máximo de k que garantice una potencia dada. El valor de k puede hallarse por ensayo y error conociendo los demás valores de la ecuación para el parámetro ϕ. Debe tenerse en cuenta que la potencia disminuye al aumentar el número de tratamientos.

7. En el modelo de componentes de varianza, tanto el número de tratamientos k como el número de repeticiones r es variable, y sus estimaciones están ligadas con el control de dichas varianzas. Un criterio de uso frecuente para elegir los valores de r y k es el de minimizar costos en la estimación de la media μ. Una medida de la cantidad de información disponible para estimar μ es la varianza de la media muestral dada por:

$$V(\bar{y}) = \frac{\hat{\sigma}^2}{rk} + \frac{\hat{\sigma}_A^2}{k}.$$

El problema se reduce a encontrar los valores de r y k que minimicen una función de costos dada por $C = c_1k + c_2kr$ para una varianza $V(\bar{y})$ fija, donde c_1 es el costo por unidad de tratamiento y c_2 es el costo por unidad experimental. La solución matemática es, Mendenhall (1968):

$$k = \frac{1}{V(\bar{y})}\left[\hat{\sigma}_A^2 + \sqrt{\hat{\sigma}_A^2\hat{\sigma}^2 c_2 / c_1}\right],\ r = \sqrt{\hat{\sigma}^2 c_1 / \hat{\sigma}_A^2 c_2}.$$

Con los datos del ejemplo 4.3, y suponiendo una varianza máxima $V(\bar{y})$, $c_1 = \$500.000$ y $c_2 = \$100.000$, se encuentra que:

$$k = (1/0.001)\left(0.0045 + \sqrt{0.0045 \times 0.0142 \times 100.000/500.000}\right) = 8.07,$$

$$r = \sqrt{0.0142 \times 500.000/0.0045 \times 100.000} = 3.97.$$

En un experimento similar al del ejemplo 4.3, y para una varianza de la media muestral no mayor de 0.001, deberían seleccionarse ocho toros y cuatro terneros por cada toro, suponiendo que los costos experimentales son de \$500.000 por cada toro y de \$100.000 por cada ternero.

Ejercicios

Observación: los resultados de estos ejercicios deben conservarse para usarlos en capítulos posteriores.

4.1 Se diseñó un experimento para estimar el número óptimo de días que ciertas vacas lecheras deberían recibir alimento ad libitum para maximizar la persistencia en la producción de leche; la cual está dada como la razón de los kilogramos de leche por día desde los 45 hasta los 270 días de lactancia, a los kilogramos de leche por día, durante los 45 primeros días de lactancia (Gill, 1978). Los datos se muestran a continuación.

Días de alimentación ad libitum

45	90	135	180	225	270
1.39	1.19	1.10	1.07	1.03	1.03
1.19	0.91	1.18	1.25	1.40	1.12
1.01	1.15	1.14	0.85	0.97	1.17
0.99	1.05	1.34	1.23	1.20	1.36

a. Describir este experimento como un proyecto de investigación. Escribir objetivos, población, factores, niveles, unidades experimentales, variable de respuesta, tamaño del experimento y explicar en qué consisten las repeticiones.
b. Describir una forma de aleatorizar el experimento.
c. Citar dos condiciones del experimento que ayuden al control del error experimental.
d. Caracterizar completamente este diseño.
e. Formular las hipótesis que se pondrán a prueba.

f. Construir la tabla del ANOVA correspondiente.
g. Estimar medias y varianzas para los tratamientos.
h. Estimar la media general y explicar si esta estimación tiene sentido en la práctica.
i. Comparar los índices promedio de persistencia en la producción de leche a los 45 días y a los 270 días de lactancia.

4.2 Gómez y González (1991) investigaron la pérdida de peso (en porcentaje del peso inicial) de la carne de res tipo milanesa después de cinco días de empacada en diferentes envolturas.

Icopor	Biopak	Cry-o-vac	Shopak
5.33	6.59	4.95	2.41
4.95	7.90	4.44	2.83
5.10	4.48	3.48	2.97
7.14	7.32	3.92	2.38
7.84	6.41	8.62	2.11

a. Especificar el diseño y el modelo del ANOVA para analizar estos datos.
b. Efectuar el ANOVA correspondiente al modelo e interpretar la prueba *F.*
c. Calcular un intervalo de confianza del 95% para la pérdida promedio de peso con la envoltura Shopak.
d. Comparar la pérdida promedio de peso entre Icopor y Biopak.
e. Describir los factores, los niveles, las unidades experimentales, la variable de respuesta y una aleatorización correcta para este diseño.
f. Calcular el número de repeticiones para un experimento similar, suponiendo una potencia de 0.90.

4.3 Madrigal y Serna (1978) compararon la población de aves en cuatro áreas de bosques diferentes: un rodal de ciprés, RC, un bosque secundario al noroeste, BSN, una plantación de pinos pátula, PP, y un bosque secundario aislado, BSA, localizadas en Piedras Blancas, departamento de Antioquia, Colombia. El total de especies observadas por día, durante diez días, fue el siguiente:

RC:	4	2	5	2	2	1	2	3	2	4
BSN:	10	10	12	11	10	12	14	12	14	11
PP:	1	1	2	3	1	1	–	2	2	3
BSA:	8	9	9	5	7	8	7	4	12	9

a. Especificar el modelo y construir la tabla del ANOVA.
b. Estimar las medias para cada uno de los bosques.

4.4 Giraldo y Vargas (1991) investigaron la morfometría de J. lamprotaenia (sardina azulita) en las Islas del Rosario, departamento de Bolívar, Colombia. Entre otras características, midieron la longitud estándar de sardinas hembras en tres estaciones de muestreo: fondo arenoso, manglar y pasto marino. Los

siguientes datos (en cm) representan parte de los datos obtenidos por las investigadoras.

Fondo arenoso	Manglar	*Pasto marino*
4.09	3.80	3.78
4.14	3.96	4.19
3.85	4.05	4.27
4.04	3.47	3.65
3.21	3.89	3.99
4.27	3.46	3.61
3.81	3.94	3.55
4.02	3.58	3.41
3.08	4.03	3.82
4.72	3.68	4.11
4.03	3.76	3.60
4.07	3.91	3.90
3.83	3.32	3.99
	4.21	3.94
	3.86	3.35
	4.02	

a. Establecer el modelo matemático y efectuar el ANOVA correspondiente para estos datos.

b. Calcular un intervalo de confianza del 95% para la diferencia de medias entre fondo arenoso y manglar.

4.5 Se midió la concentración de calcio en hojas de una especie de árbol. Se tomaron al azar cuatro hojas y se hicieron cuatro lecturas de cada hoja, lográndose los siguientes resultados en porcentaje de peso seco (Snedecor y Cochran, 1984)

Hojas	Lecturas			
1	3.28	3.09	3.03	3.03
2	3.52	3.48	3.38	3.38
3	2.88	2.80	2.81	2.76
4	3.34	3.38	3.23	3.26

a. Describir de forma completa el experimento especificando objetivos, población, factores, tratamientos, unidades experimentales y variable de respuesta.

b. Indicar un método de aleatorización apropiado para el experimento.

c. Establecer el modelo del ANOVA para analizar estos datos y construir la tabla correspondiente.

d. Enunciar las hipótesis estadísticas del problema.

e. Estimar los porcentajes de variación entre hojas y entre lecturas.

f. Calcular un intervalo del 95% de confianza para la concentración promedio del calcio por hoja.

g. Estimar el número de repeticiones necesarias para alcanzar una po-

tencia del 80%. Discutir la conveniencia de aumentar hojas en vez de repeticiones o lecturas.

4.6 Se realizó un estudio piloto para determinar la variabilidad en la rata de sedimentación de eritrocitos en infantes de un día de nacidos. De cada niño, elegido al azar entre los recién nacidos, se tomaron varias muestras que se entregaron a diferentes laboratoristas para su análisis. Se tienen los siguientes resultados.

Muestras	Infantes							
	1	2	3	4	5	6	7	8
1	110	130	50	130	45	120	100	130
2	90	115	75	45	55	50	200	80
3	120	105	85	50	65	150	90	70
4			40	40			70	80
5							90	150

a. Construir la tabla del ANOVA y analizarla.
b. Calcular un intervalo del 95% de confiabilidad para la varianza entre infantes. Discutir el intervalo obtenido.
c. Estimar la variabilidad entre laboratoristas y dar el porcentaje sobre la variabilidad total.

4.7 Ortega (1992) estudió los hábitos alimenticios de M. Ocellata como depredador de larvas de C. quinquefasciatus. En particular, comparó el consumo promedio diario de larvas en los últimos cinco instares del depredador. Una parte de los resultados obtenidos es la siguiente:

Instares				
7	8	9	10	11
25.16	24.22	27.56	25.58	24.20
20.85	27.75	27.10	23.67	23.95
20.00	25.14	26.64	24.73	25.65
20.67	29.25	26.63	25.30	24.95
18.67	25.25	29.38	22.75	25.65
20.33	27.40	28.00	24.18	24.26
19.00	24.60	28.25	26.50	25.38
21.33	25.60	27.67	25.36	24.94
17.00		26.90	27.08	23.61
21.50		27.56	23.42	23.72
		28.89	24.85	22.71
		24.10	21.09	
		25.56		
		28.33		
		27.50		

a. Efectuar el ANOVA para estos datos.

b. Estimar el consumo promedio en cada instar.
c. Comparar el consumo promedio de los instares séptimo y undécimo.

4.8 Un grupo de ratones con carcinomas mamarios fue tratado por irradiación. Después del tratamiento se tomaron biopsias del tumor y se realizaron varios recuentos mitóticos, CPM, con los siguientes resultados: (Gill, 1978).

Conteos	Ratones									
	1	2	3	4	5	6	7	8	9	10
1	21	68	50	11	1	26	12	47	36	31
2	19	73	52	10	12	15	12	47	31	27
3	26	70	59	12	11	11	9	40	42	24
4		60	48	11			15	42		

a. Establecer el modelo y realizar el ANOVA correspondiente.
b. Estimar las componentes de varianza.
c. Al repetir el experimento cambiando la dosis de irradiación, qué se aconseja: ¿aumentar unidades experimentales o recuentos mitóticos? Diferenciar las fuentes de variación y explicar el origen de cada una de ellas.

4.9 Los islotes beta de Langerhans son las células productoras de insulina en el páncreas. Los datos siguientes representan la producción de insulina por islotes del páncreas de ratas obesas, en un periodo de tres semanas. Inicialmente se prepararon 36 cultivos y semanalmente se tomaron nueve para calcular la cantidad de insulina producida (Koopmans, 1987).

Semanas			
0	1	2	3
31.2	18.4	55.2	69.2
72.0	37.2	70.4	52.0
31.2	24.0	40.0	42.8
28.2	20.0	42.8	40.6
26.4	20.6	26.8	31.6
40.2	32.2	80.4	66.8
27.2	23.0	60.4	62.0
33.4	22.2	65.6	59.2
17.6	7.8	15.8	22.4

a. Efectuar el ANOVA para estos datos. ¿A qué conclusión puede llegarse?
b. Mirar las cajas esquemáticas para cada semana y analizar variabilidad y tendencia en la producción de insulina con el tiempo.

4.10 Se conjetura que el peso de un niño al nacer está asociado con el nivel de consumo de cigarrillos de la madre durante el primer trimestre del embarazo. Un estudio realizado en una clínica de maternidad dio los siguientes resultados, para madres agrupadas según el hábito de fumar. Los datos son los pesos en g de los niños al nacer (Rosner, 1999).

No fumadoras	Ex fumadoras	Fumadoras (menos de un paquete)	Fumadoras (más de un paquete)
3400	2620	2670	2800
2800	3300	2800	3080
3120	3710	2630	2580
3350	3210	2130	2220
4170	3530	3760	2800
3760		3260	3200
3440		2800	2630
			2450

a. Efectuar el ANOVA e interpretar la prueba F.
b. Calcular medias y varianzas para cada grupo.
c. Comparar el peso promedio en no fumadoras y en ex fumadoras.

5

Supuestos y alternativas del análisis de varianza

5.1 Supuestos

La validez del análisis de varianza está supeditada al cumplimiento de algunos requisitos de tipo matemático y probabilístico en los datos u observaciones. Se dice que el modelo del ANOVA es apto para comparar efectos de factores en un fenómeno si los datos experimentales se ajustan al modelo, es decir, si las condiciones subyacentes al modelo están implícitas en los datos. Estas condiciones, que en adelante llamaremos supuestos, fueron mencionadas en el capítulo anterior y serán el tema de análisis del capítulo presente. Una referencia fundamental y obligada en el estudio de los supuestos del ANOVA la conforman los tres artículos publicados en *Biometrics* por Eisenhart, Cochran y Bartlett (1947), respectivamente.

Los supuestos para los modelos I y II pueden resumirse en:

1. Independencia de los errores ε_{ij}.
2. Distribución Normal de los errores ε_{ij}.
3. Homogeneidad de las varianzas de los tratamientos.
4. Aditividad o linealidad en los parámetros del modelo; aplicable, en particular, a los modelos factoriales que se verán posteriormente.

El examen de estos supuestos incluye los métodos para probar su cumplimiento, las consecuencias que se derivan de su no cumplimiento y los pasos que se deben seguir para remediar estas consecuencias.

Normalidad de los errores

La normalidad de los errores debe probarse para cada uno de los *k* grupos o subpoblaciones, pero a menos que los grupos sean muy grandes, es suficiente la prueba de normalidad para los errores ε_{ij} en conjunto, sobre todo si los grupos tienen varianzas similares.

De otro lado, la *robustez* de una prueba estadística se refiere a la propiedad de

resultar poco afectada cuando no se satisface completamente algún supuesto de validez de la prueba. La prueba F es robusta para el error α en el caso de no normalidad. Para el modelo I, la falta de normalidad en los datos afecta muy poco la significancia de la prueba F, siempre que la desviación no sea extrema. Cochran (1947) ha conjeturado que la verdadera probabilidad correspondiente a un error $\alpha = 0.05$ puede estar entre 0.04 y 0.07, mientras que para $\alpha = 0.01$, entre 0.005 y 0.020.

En general, los intervalos de confianza y los contrastes entre medias resultan más afectados, sobre todo cuando los tamaños r_j son pequeños. En el modelo II, la falta de normalidad tiene efectos mayores y, en particular, los intervalos de confianza para las varianzas pueden resultar seriamente afectados.

Los errores ε_i son valores poblacionales desconocidos, pero pueden estimarse mediante las diferencias $e_{ij} = y_{ij} - \bar{y}_{.j}$. Los valores e_{ij} se llaman residuos y es a estas diferencias a las que se les aplica la prueba de normalidad. Dicha prueba puede ser cualquiera de las enumeradas en la sección 3.7, u otra que haya sido validada estadísticamente.

Los residuos para el ejemplo 4.1 son los siguientes:

y_{ij} :	41	44	45	43	42	48	49	49	49	45	40
e_{ij} :	−2.0	1.0	2.0	0.0	−1.0	0.0	1.0	1.0	1.0	−3.0	−6.4
y_{ij} :	50	44	48	50	40	39	46	46	41		
e_{ij} :	3.6	−2.4	1.6	3.6	−2.4	−3.4	3.6	3.6	−1.4		

La prueba de ajuste a la normal de Kolmogorov-Smirnov dice que estos residuos tienen distribución Normal ($p = 0.81$). De acuerdo con la gráfica de probabilidad Normal (véase figura 5.1), se puede concluir, igualmente, que el ajuste a la normal es bueno.

La misma prueba de Kolmogorov-Smirnov da una probabilidad $p = 0.85$ para los residuos del ejemplo 4.2 y $p = 0.82$ para los del ejemplo 4.3. Las gráficas de normalidad, figuras 5.2 y 5.3 respectivamente, indican buen ajuste.

Homogeneidad de las varianzas

Uno de los supuestos más importantes y cruciales en el ANOVA, es el de igualdad de varianzas entre los grupos que se están comparando. La prueba correspondiente se conoce como prueba de homogeneidad de varianzas o prueba de homoscedasticidad. Si las varianzas son diferentes, se dice que hay heteroscedasticidad.

La prueba F del ANOVA modelo I es robusta y trabaja bien aun con varianzas considerablemente heterogéneas, sobre todo si los tamaños r_j son similares. Box (1954) demostró que para desviaciones relativamente grandes —una varianza hasta nueve veces mayor que otra— el nivel α puede variar solamente de 0.05 a aproximadamente 0.06, cambio que tiene poca importancia en la práctica. Cuando los tamaños de los grupos son bastante diferentes entre sí, el error de tipo I se aleja notablemente del nivel de significación α; si las varianzas mayores están asociadas con los tamaños de muestras más grandes, el error de tipo I será menor que α, pero

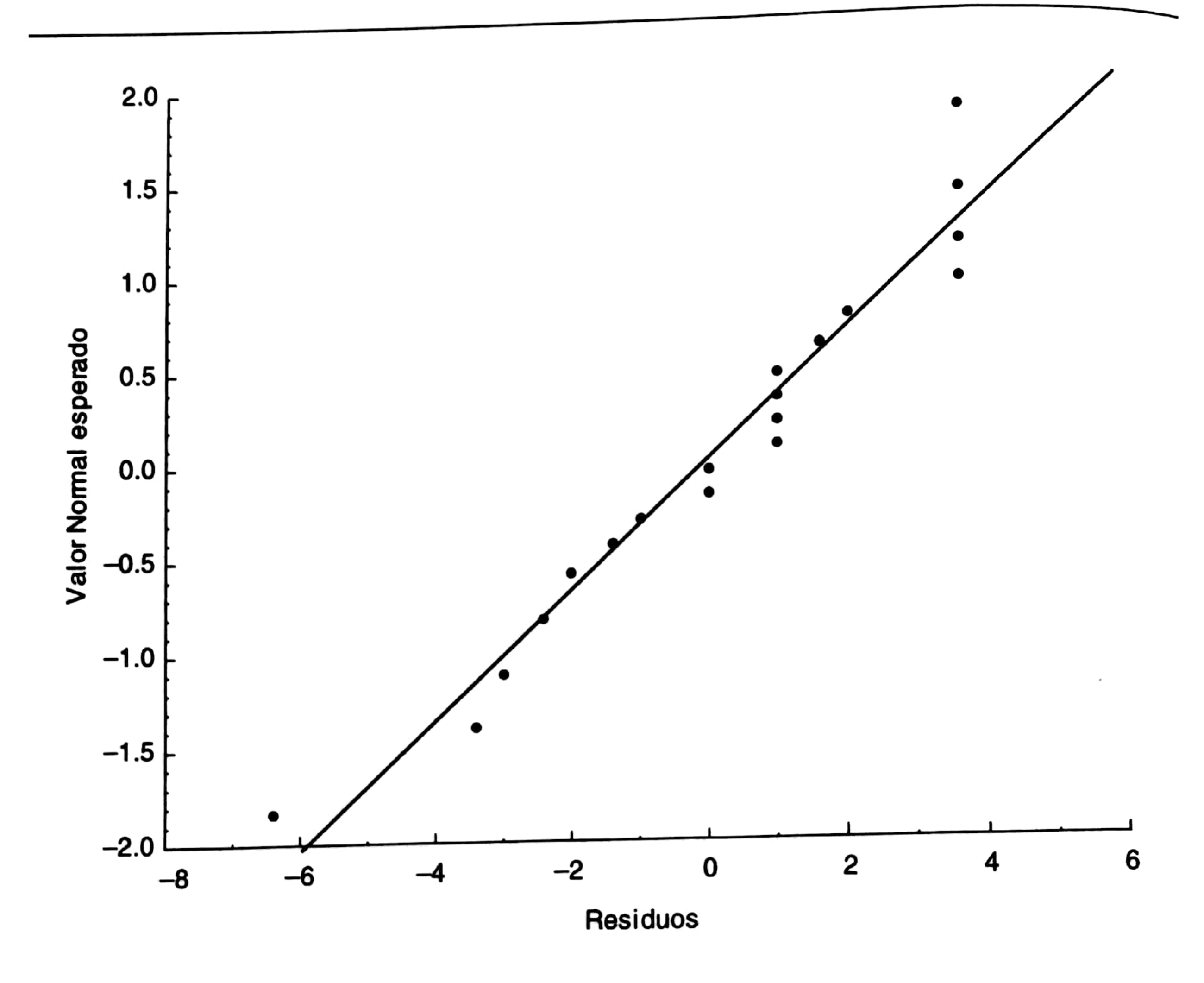


Figura 5.1 Normalidad para residuos, ejemplo 4.1

si están asociadas con las muestras pequeñas, entonces la probabilidad de error será mayor que α (Zar, 1999). Por otra parte, las comparaciones de medias y los intervalos de confianza resultan afectados cuando las varianzas son heterogéneas.

En el modelo II, la desigualdad de varianzas genera efectos adversos en las estimaciones de las componentes de varianza, aun para grupos iguales.

Para probar la hipótesis de homoscedasticidad existen varios procedimientos; uno de los más conocidos es la prueba desarrollada por Bartlett y publicada en 1937. Las hipótesis por probar son:

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_\kappa^2$; es decir, las varianzas son homogéneas.

H_1: no todas las varianzas son iguales.

La regla de decisión de esta prueba consiste en rechazar H_0 si M/C > $\chi^2_{\alpha, k-1}$. Los valores de M y de C se calculan a partir de la tabla 5.1 como sigue:

$$M = \left[\nu \ln \bar{s}^2 - \sum_j \nu_j \ln s_j^2 \right]; \text{ con } \bar{s}^2 = \sum_j \nu_j s_j^2 / \nu.$$

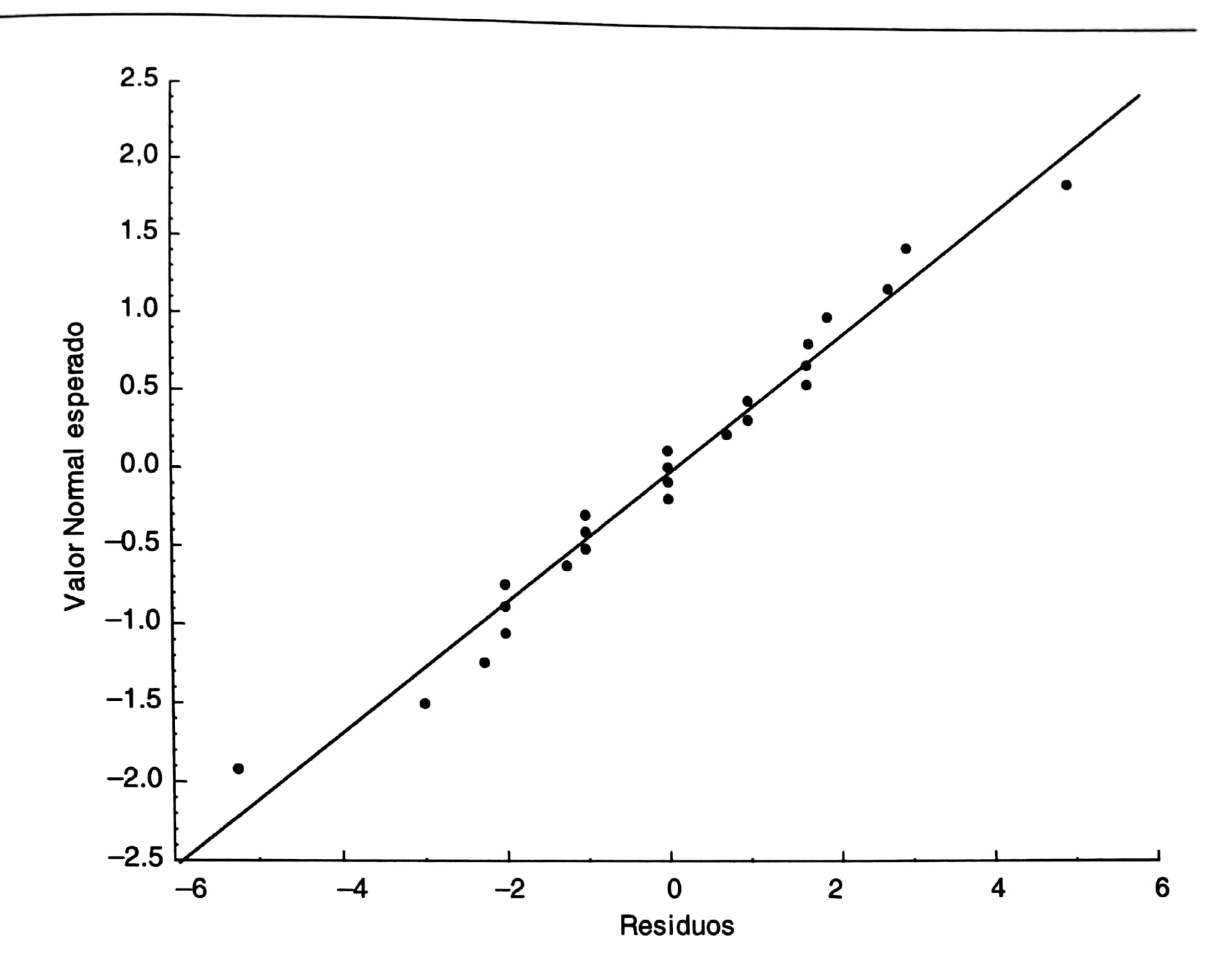


Figura 5.2 Normalidad para residuos, ejemplo 4.2

$$C = 1 + \frac{1}{3(k-1)}\left(\sum_j \frac{1}{v_j} - \frac{1}{v}\right); \text{ con } v = \sum_j v_j \text{ y } v_j = r_j - 1.$$

Ejemplo 5.1 Aplicar la prueba de Bartlett para chequear la homogeneidad de varianzas en el ejemplo 4.2. Los datos se organizan como en la tabla 5.1.

El promedio de las varianzas es $\overline{s}^2 = 112.0 / 20 = 5.6$.

$$M = (20 \times 1.7228 - 32.6366) = 1.8187,$$
$$C = 1 + (0.8759 - 1/20)/9 = 1.092,$$
$$M/C = 1.665,\ \chi^2_{0.05,3} = 7.81, p = 0.725.$$

El factor de corrección C siempre es mayor que la unidad, de modo que debe evaluarse solamente cuando M está cercano al valor crítico calculado.

La prueba de Bartlett puede aplicarse a grupos iguales o desiguales, tiene buena potencia, pero desgraciadamente es muy sensible a la no normalidad de los datos. Un resultado significativo de esta prueba podría indicar falta de normalidad en vez de heterogeneidad en las varianzas; también exige independencia de los errores ε_{ij}. Algunos autores no la recomiendan para tamaños de muestra $r_j < 4$.

La prueba de Hartley para homogeneidad de varianzas involucra el cociente entre las varianzas mayor y menor; es simple de calcular pero exige tablas especiales

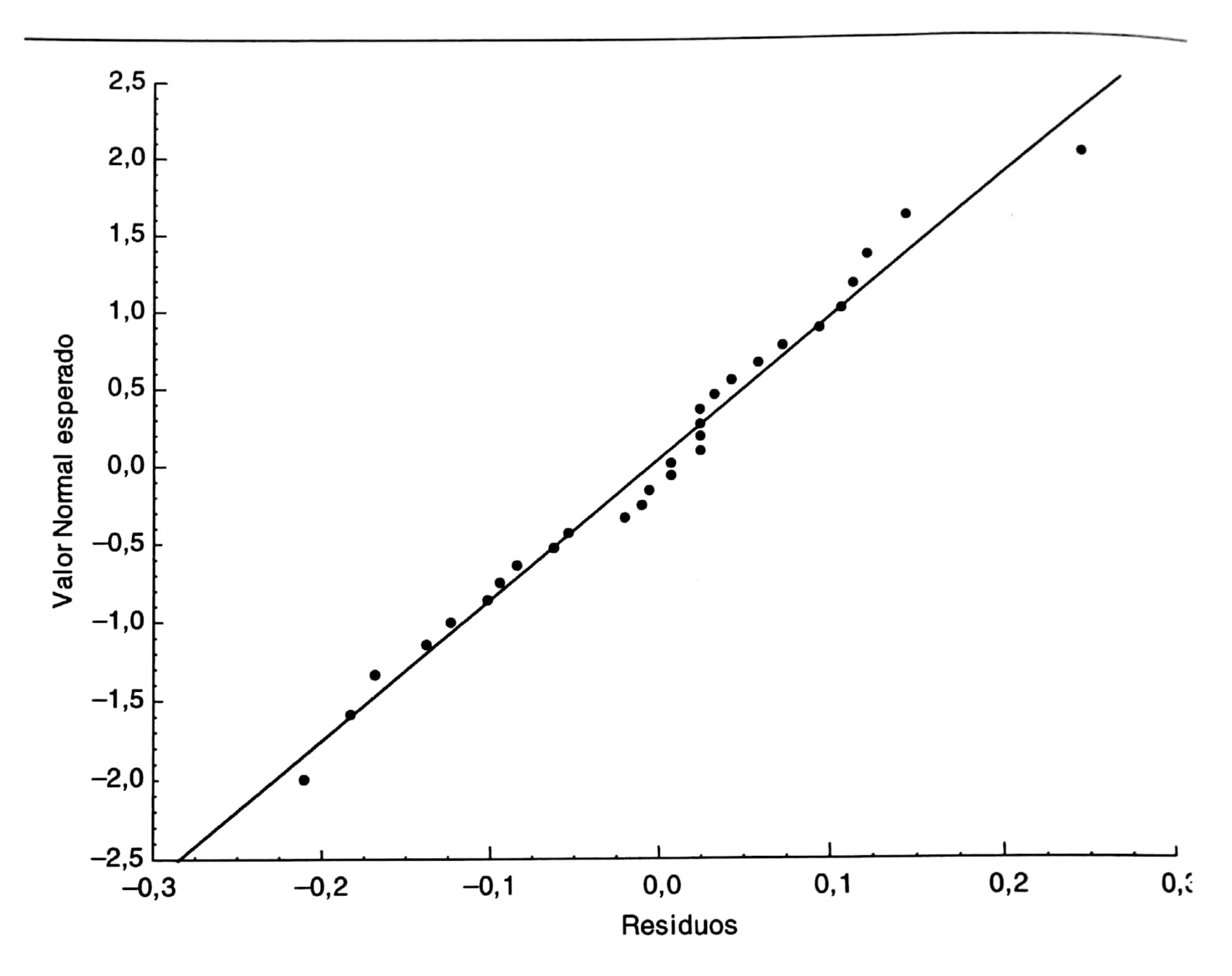


Figura 5.3 Normalidad para residuos, ejemplo 4.3

Tabla 5.1 Cálculos para la prueba de Bartlett aplicada a los datos del ejemplo 4.2

Dietas	υ_j	s_j^2	$\nu_j s_j^2$	$\ln s_j^2$	$\upsilon_j \ln s_j^2$	$1 / \upsilon_j$
D_1	3	3.33	10.0	1.2039	3.6117	0.333
D_2	5	8.00	40.0	2.0794	10.3970	0.200
D_3	5	2.80	14.0	1.0296	5.1480	0.200
D_4	7	6.86	48.0	1.9257	13.4799	0.143
Totales	20		112.0	6.2386	32.6366	0.876

para el contraste, disponibles solo para grupos iguales. Esta prueba también exige normalidad en los datos e independencia en los errores ε_{ij}.

Si el cociente $F_{\text{máx}} = \text{máx}\, s_j^2 / \text{mín}\, s_j^2$ es mayor que un valor crítico tabulado $f_{\alpha,k,r-1}$, se rechaza H_0. Para repeticiones levemente desbalanceadas, la prueba es satisfactoria si se reemplazan los grados de libertad $r - 1$ por $[(1/k)\Sigma r_j] - 1$.

Si se toma de nuevo el ejemplo 4.2, se encuentra que $F = 8.0 / 2.8 = 2.86$. El valor crítico $f_{0.05,4,5} = 13.7$ (véase tabla A.8), indica que no hay evidencia de heterogeneidad en las varianzas.

Existen, en la literatura estadística, muchas pruebas de homogeneidad de varianzas con los mismos nombres de sus autores, ejemplos: Cochran, Levene, O'Brien,

Brown-Forsythe y Burr-Foster. Se describe a continuación esta última ya que admite grupos pequeños y desiguales, es simple de aplicar, es robusta a la carencia de normalidad y varianzas iguales a cero no interrumpen la prueba como ocurre en la de Bartlett. Igualmente, se recomienda la prueba Levene para muestras aleatorias de distribuciones continuas ya que no exige normalidad y se encuentra en varios paquetes estadísticos de uso rutinario.

La prueba Q de Burr-Foster calcula una estadística q dada por el cociente:

$$q = \bar{\upsilon}\left(\upsilon_1 s_1^4 + \cdots + \upsilon_k s_k^4\right) \Big/ \left(\upsilon_1 s_1^2 + \cdots + \upsilon_k s_k^2\right)^2.$$

Valores grandes de q implican el rechazo de la hipótesis de varianzas poblacionales iguales. El valor crítico $q_{a,k,\upsilon}$ se encuentra tabulado para algunos valores de $1 - \alpha$, de k y de υ ($\upsilon = r - 1$) en el apéndice, tabla A.9.

Para grupos desiguales puede usarse el promedio $\bar{\upsilon} = (1/k)\sum \upsilon_j$; en vez de υ, además se puede interpolar el valor de $\bar{\upsilon}$ si es necesario.

Para $r = 2$, la estadística q se convierte en $\Sigma R_j^4 / (\Sigma R_j^2)^2$, donde R_j, es el rango o diferencia entre los dos datos.

Ejemplo 5.2 Para el ejemplo 4.1, la prueba de Burr-Foster (véase tabla 5.2) se aplica como sigue:

La estadística de prueba es $q = 496.38 / (35.6)^2 = 0.392$, grupos iguales. Con $\upsilon = 4$ y $k = 4$ se lee $q_{0.99,4,4} = 0.549$, lo que permite concluir que hay homogeneidad en las varianzas poblacionales. Para esta prueba se considera un $\alpha = 0.01$.

La prueba F del ANOVA es robusta para grupos iguales, de modo que la mejor estrategia contra efectos negativos de la heterogeneidad de varianzas es el diseño balanceado; es decir, igual número de réplicas por celda. Si la heterogeneidad es muy pronunciada, una alternativa que puede seguirse es la transformación de los datos a otra escala, con lo cual se busca estabilizar las varianzas.

Para transformar los datos en la prueba de homogeneidad, o más concretamente, para decidir a partir de qué valor de α deberán transformarse los datos, Anderson y McLean (1974) sugieren las reglas siguientes:

1. Si se acepta homogeneidad al nivel $\alpha = 0.01$, no es necesario transformar los datos.

2. Si se rechaza la hipótesis de homogeneidad con $p < 0.001$, una alternativa es transformar los datos.

3. Si la significancia de la prueba está entre 0.01 y 0.001, deberá averiguarse con insistencia la distribución teórica de los datos o una razón práctica para transformarlos. De no encontrarla, es preferible no hacer transformaciones.

Tabla 5.2 Datos para la prueba de Burr-Foster

Varianzas	*Tratamientos*				
	T_1	T_2	T_3	T_4	*Total*
s_j^2	2.5	3.0	18.8	11.3	35.6
s_j^4	6.25	9.0	353.44	127.69	496.38

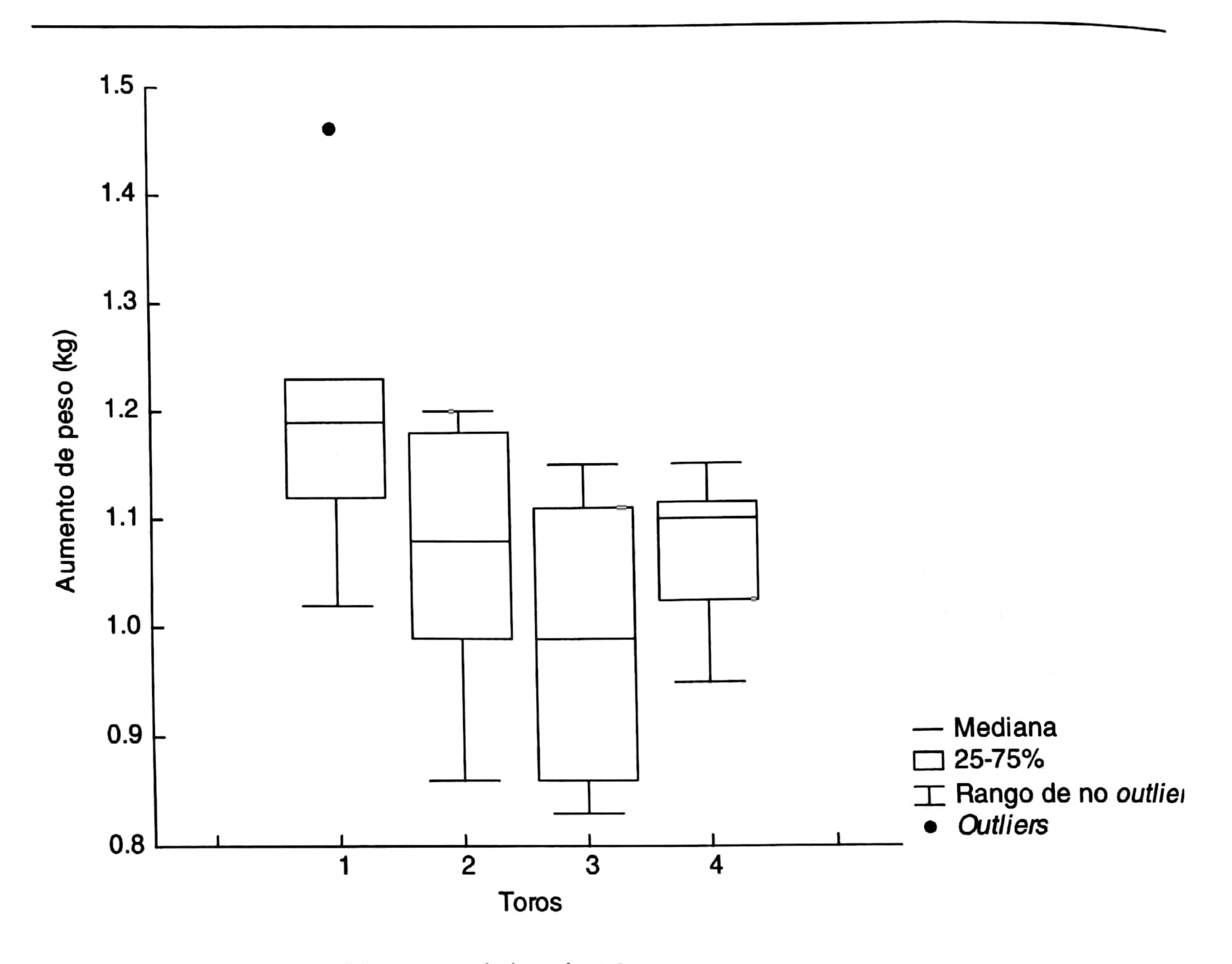


Figura 5.4 Cajas esquemáticas para el ejemplo 4.3

Si se usa la prueba de Bartlett y el investigador no tiene conocimiento teórico de su variable, se sugiere usar $\alpha = 0.001$, sobre todo cuando la distribución de los datos presenta colas demasiado largas. Esta sugerencia previene al analista contra transformaciones innecesarias.

En muchas investigaciones, los resultados cumplen los supuestos cabalmente, con lo cual es suficiente un chequeo rápido mediante los gráficos que ofrecen algunos programas de computador. El gráfico que mejor describe la variabilidad y permite su comparación es el de las cajas esquemáticas. La figura 5.4 muestra las cajas esquemáticas para los grupos del ejemplo 4.3. Se nota que el grupo 1 tiene un dato bastante extraño y, posiblemente, por eso su varianza es similar a la de los otros grupos.

Independencia de los errores

El supuesto de independencia en los errores ε_{ij} es definitivo para el ANOVA. La falta de independencia afecta la validez de la prueba F como también las estimaciones, tanto para el modelo I como para el modelo II. La dependencia en los errores se caracteriza porque en los residuos aparecen secuencias largas de valores positivos o negativos, o de ambos, o resultan series en las cuales se alternan un valor positivo con uno negativo. En experimentación agrícola, por ejemplo, se sabe que las parce-

las adyacentes producen cantidades similares entre sí y diferentes de las de parcelas más apartadas.

La falta de independencia puede darse también por correlación en el tiempo. Si se sospecha correlación o dependencia en los datos, puede aplicarse una prueba de tendencias o *rachas*. Gibbons (1993) y Sokal y Rohlf (1995), traen varias pruebas aplicables a residuos, conocidas como pruebas de independencia o pruebas de aleatoriedad en muestras.

No existen transformaciones o ajustes simples que corrijan la falta de independencia en los errores. La solución apropiada sería la modificación del diseño básico o la forma de realizar el experimento. A menudo un diseño de bloques ayuda a corregir este tipo de dependencia.

Cabe recordar que el mejor seguro contra la correlación o dependencia en los errores es la aleatorización inicial del experimento. La no aleatoriedad puede reflejarse en una gran divergencia de cualquiera o de todos los supuestos del ANOVA. Por otra parte, en estudios observacionales no siempre es posible la aleatorización, pero la introducción del tiempo como un factor en el modelo corrige, generalmente, dicha correlación.

5.2 Análisis gráfico de los residuos

Sin ser un método de prueba riguroso, el análisis gráfico es un procedimiento muy efectivo para descubrir los casos de verdadero alejamiento del modelo adoptado. Mediante el estudio de los residuos e_{ij} se pueden descubrir discrepancias entre el modelo teórico y los datos; tales discrepancias permanecerían encubiertas si solo se analiza la tabla del ANOVA.

Cuando se cumplen los supuestos inherentes al modelo apropiado, se espera que los residuos varíen aleatoriamente. Al contrario, si se encuentra que los residuos tienen tendencias sistemáticas no explicables, hay razón para dudar de la aptitud del modelo. Por tanto, como una acción preliminar a los análisis estadísticos programados, deberían construirse y explorarse los gráficos de residuos.

La tabla 5.3 contiene los residuos para el ejemplo 4.1, y entre paréntesis el orden, en el tiempo, en que fueron obtenidos los datos de acuerdo con la tabla 4.1.

Los residuos pueden graficarse de distintas maneras y hoy día existen programas de computador que traen diferentes opciones. No obstante, debe tenerse claridad

Tabla 5.3 Residuos y orden de las observaciones en el tiempo, para los datos del ejemplo 4.1

	T_1	T_2	T_3	T_4
	−2 (7)	0 (1)	−6.4 (4)	−2.4 (5)
	1 (11)	1 (1)	3.6 (10)	−3.4 (8)
	2 (12)	1 (3)	−2.4 (13)	3.6 (9)
	0 (17)	1 (6)	1.6 (15)	3.6 (19)
	−1 (18)	−3 (14)	3.6 (16)	−1.4 (20)
$\bar{y}_{.j}$	43.0	48.0	46.4	42.4

sobre el gráfico que se desea, pues de nada sirve un gráfico muy bonito si no va acompañado de la interpretación correcta. Algunas características que se pueden examinar en los gráficos de residuos, se describen a continuación.

Normalidad

Los gráficos de probabilidad normal para muestras o residuos vistos en secciones anteriores, ayudan a reconocer el alejamiento de la normalidad. Otro gráfico que también refleja el grado de normalidad es el diagrama de puntos construido con los residuos (véase figura 5.5). Si se tienen pocos datos, el gráfico resulta bastante ambiguo, ya que pueden presentarse fluctuaciones con apariencia de no normalidad. De otro lado, con un número mayor de datos puede tenerse una buena imagen de una distribución Normal centrada en el punto cero.

El alejamiento de la normalidad lo dan, generalmente, los valores extremos, muy grandes o muy pequeños comparados con el resto, o la acumulación de puntos lejos del punto cero.

Homogeneidad de las varianzas

Si el modelo matemático es apropiado, los residuos no estarán relacionados con los valores de la variable de respuesta. En particular, el gráfico de residuos contra los niveles del factor deberá mostrar un comportamiento similar para todos los niveles, como en la figura 5.6 que corresponde a los residuos del ejemplo 4.1. Algunas veces la varianza aumenta con la media o una aumenta y la otra disminuye. Esta variación se detecta por la forma de *embudo* que toma la gráfica (véase figura 5.7).

Independencia

En muchos experimentos no es posible mantener constantes las condiciones durante todo el proceso. Pueden cambiar los factores ambientales, mejorar la habilidad del experimentador o improvisarse aparatos, sujetos y métodos, lo que se refleja en la correlación de residuos. Si estos se grafican contra el tiempo, es decir, si se tiene en cuenta el orden en que fueron tomadas las observaciones, pueden aparecer tendencias crecientes o decrecientes como se muestra en la figura 5.8. En la figura 5.9 no hay razón para sospechar tendencias en los residuos y puede suponerse independencia entre ellos.

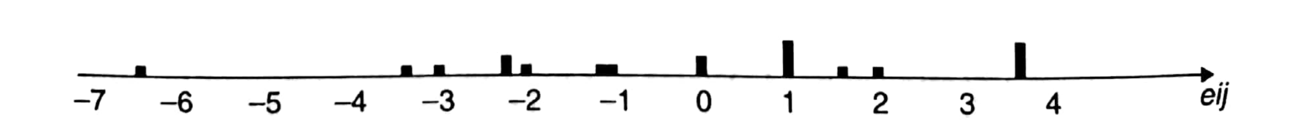


Figura 5.5 Diagrama lineal de puntos para los residuos de la tabla 5.3

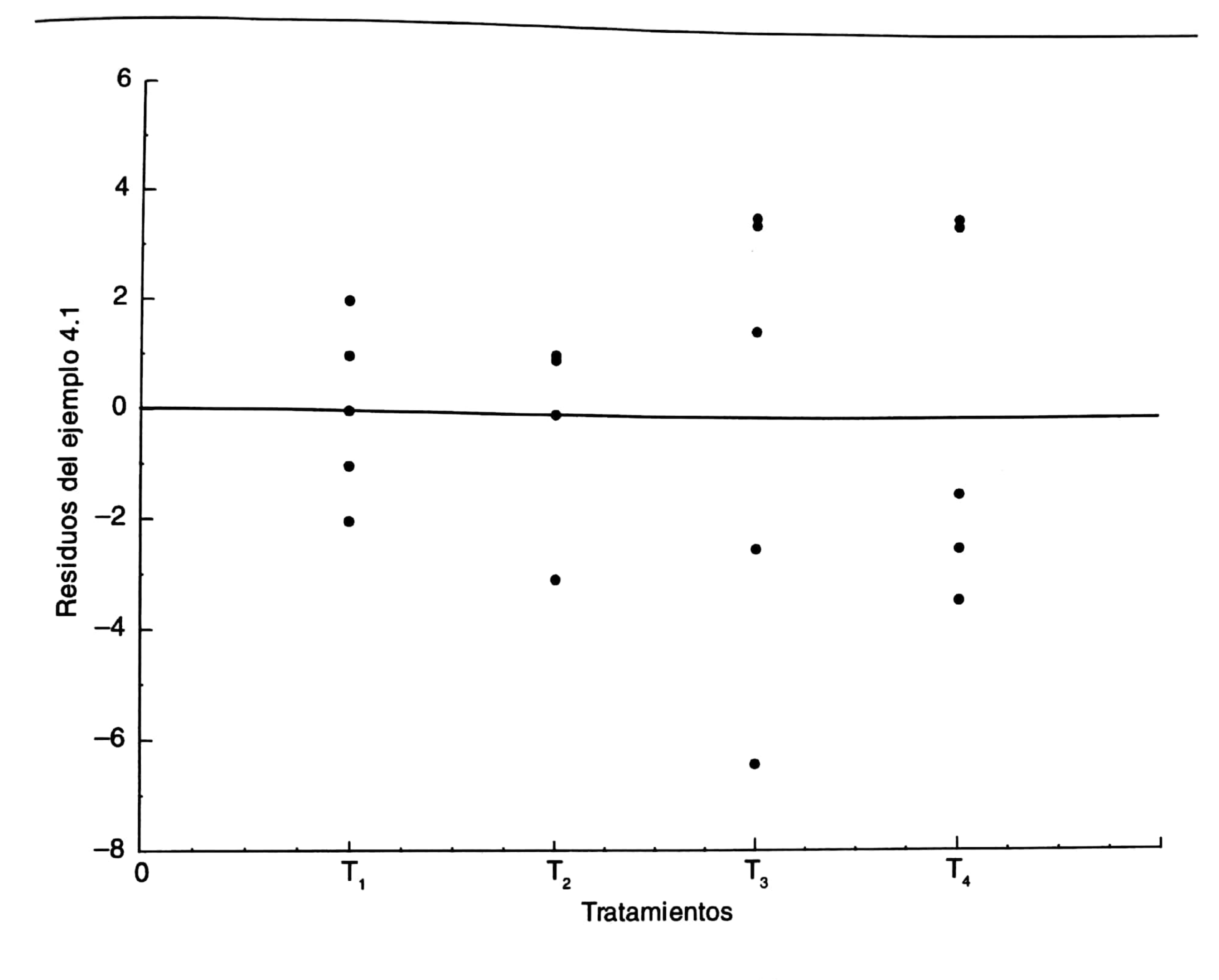


Figura 5.6 Residuos contra niveles del factor A: tratamiento médico

Valores extremos u outliers

El gráfico de residuos contra el tiempo puede mostrar que no existe violación de los supuestos —cuando los puntos están distribuidos al azar a lado y lado de la línea cero formando una especie de banda rectangular más o menos uniforme—, pero aparecen uno o más valores por fuera de lo común, los cuales hacen pensar en una posible falla en los registros originales. Como ya se dijo, este dato debe investigarse cuidadosamente antes de encontrar una débil excusa para un probable rechazo. Para ello, existe gran variedad de pruebas; muchas de estas exigen tablas extensas no siempre disponibles. El texto de Barnett y Lewis (1994), por ejemplo, es amplio en tablas y contiene extensa bibliografía sobre el tema.

Omisión de variables

Los residuos pueden graficarse contra cualquier variable de importancia en la investigación. Aunque no se observen comportamientos anormales que invaliden el modelo, se descubren influencias de otras variables que sería necesario controlar, o que podrían incluirse en el modelo inicial como factores por estudiar en trabajos

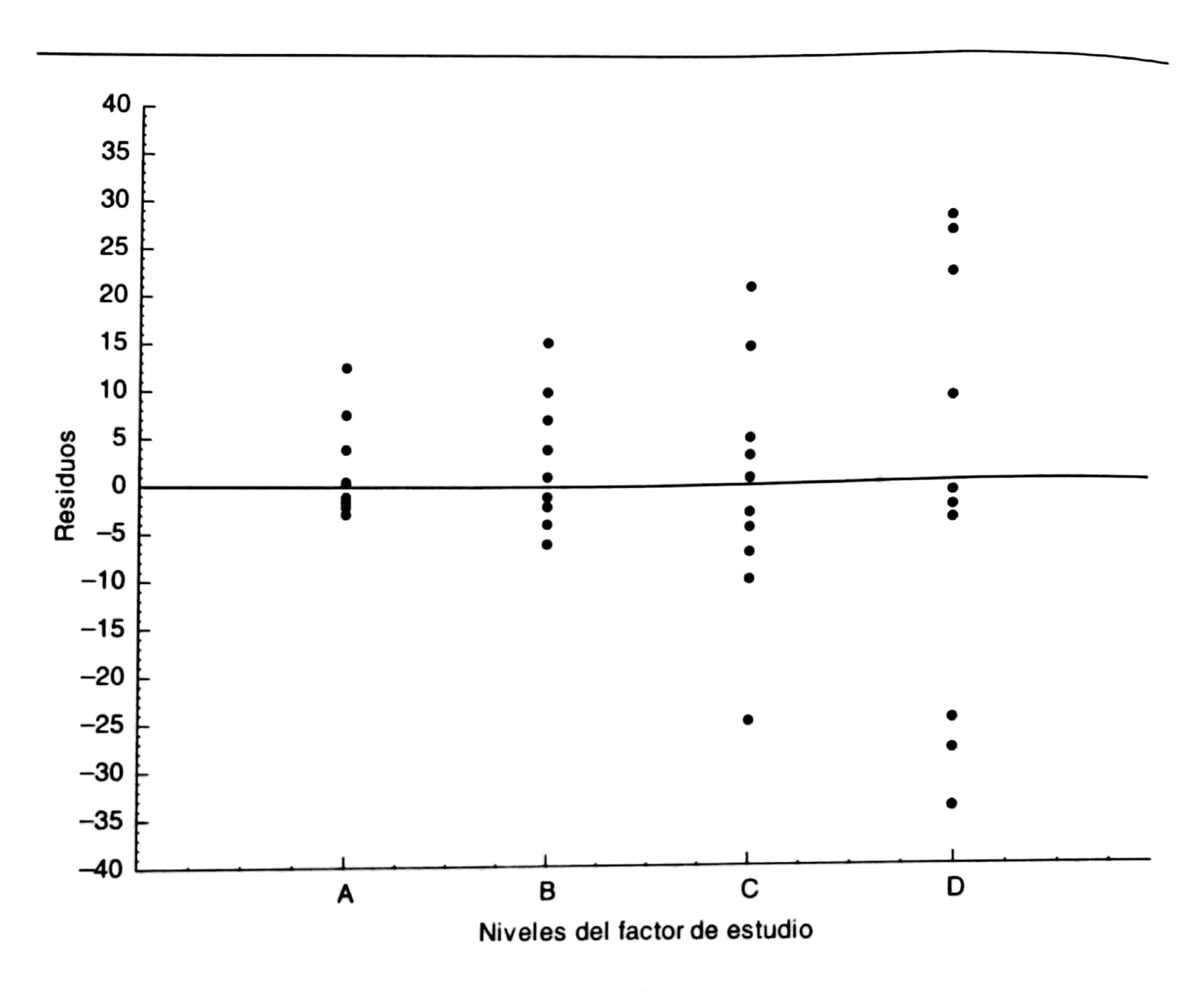


Figura 5.7 Residuos contra niveles del factor de estudio

futuros. Para el ejemplo 4.1, algunas de esas variables podrían ser el sexo, la edad o los antecedentes de la enfermedad.

Según Box, Hunter y Hunter (2002), si se pudiera confiar en que los supuestos de independencia, normalidad y homoscedasticidad fueran válidos, entonces la *MCE* y las medias $\bar{y}_{.j}$ serían, conjuntamente, estadísticas suficientes para σ^2 y las medias μ_j, únicos parámetros que tiene el modelo. Se podría decir que, una vez calculadas las estadísticas, no queda información relevante en los datos originales, y por tanto se podrían ignorar los residuos y las observaciones originales y concentrarse totalmente en la interpretación de estas estadísticas. En la práctica, sin embargo, no es aconsejable abandonar el chequeo de los residuos porque los datos pueden contener información valiosa no incluida en el modelo matemático y, por consiguiente, no revelada en la tabla del ANOVA.

5.3 Transformaciones de los datos

Si el análisis de residuos u otro análisis de los datos indica que el modelo de ANOVA no es aplicable, entonces se presentan varias opciones para elegir: transformar los datos, usar una técnica no paramétrica, hacer una reducción de los datos, por ejemplo

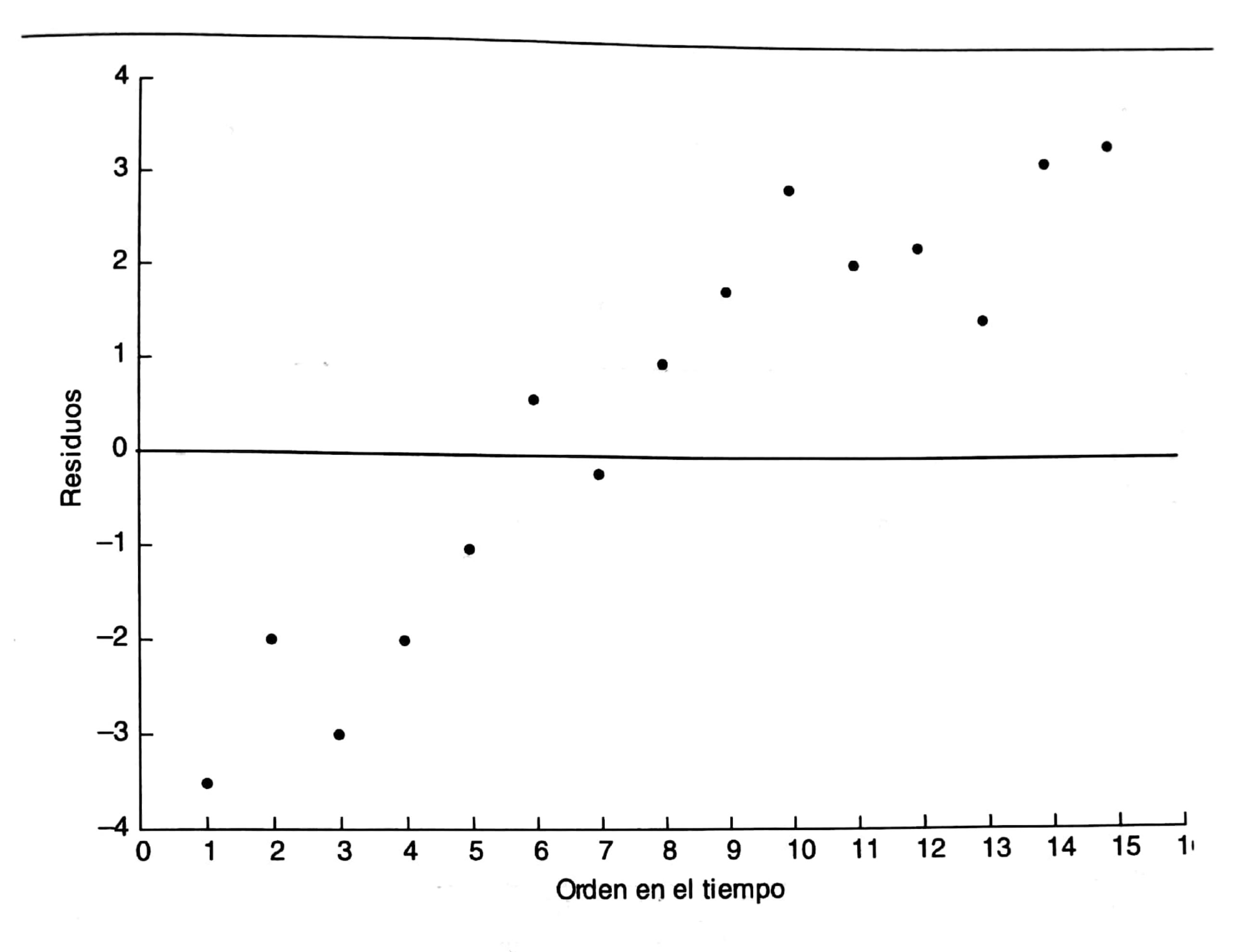


Figura 5.8 Residuos contra orden en el tiempo

medias recortadas, o recurrir a métodos aproximados. Según Weerahandi (1996), una alternativa cuando las varianzas son heterogéneas es aplicar la prueba F modificada de Fisher-Pitman. Pero, si el supuesto de normalidad es razonable, una alternativa mejor es la prueba de Welch que no requiere homoscedasticidad. Un enfoque diferente para analizar datos experimentales es el modelo lineal general, cuyos supuestos son menos restrictivos que los del ANOVA, Dobson (2001).

Una manera de resolver el problema de la heterogeneidad de las varianzas es transformando los datos originales a otra escala numérica. En muchas ocasiones, la transformación corrige simultáneamente la falta de normalidad y de aditividad de los datos.

Saber qué transformación es la más adecuada para un conjunto de datos es un problema más o menos complicado. Sin embargo, el conocimiento del comportamiento probabilístico de la variable en estudio ayuda en la elección de la transformación apropiada. A continuación se presentan algunas de las transformaciones de datos más utilizadas.

Transformación a la escala de raíz cuadrada

Cuando la heterogeneidad se debe a que las varianzas cambian con las medias, o sea, el cociente $s_j^2 / \bar{y}_{.j}$ es aproximadamente constante, la transformación a raíces

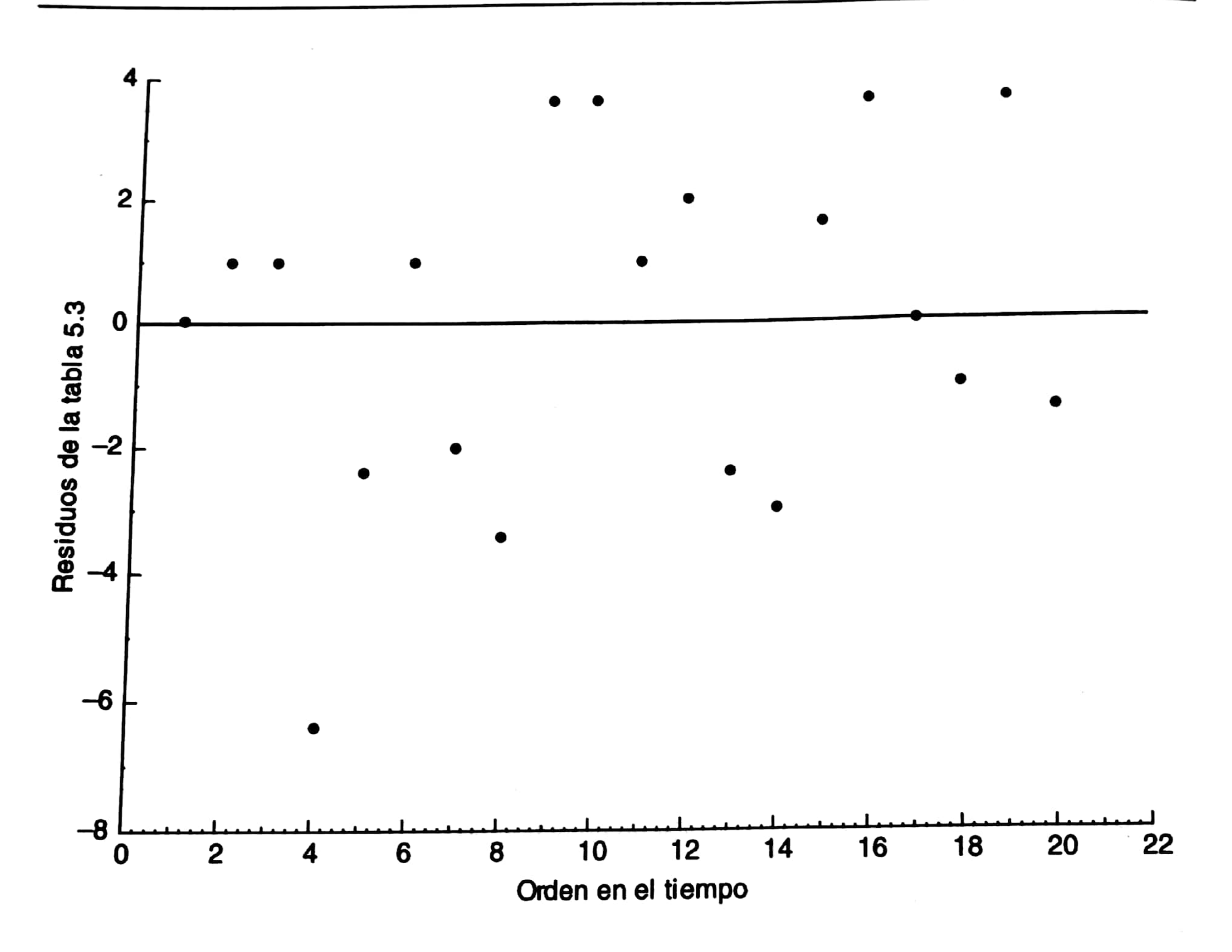


Figura 5.9 Residuos contra orden en el tiempo

cuadradas es la más apropiada para estabilizar las varianzas. Algunos autores recomiendan usar $\sqrt{Y + 0.5}$ en vez de $\sqrt{Y}$, sobre todo cuando se tienen datos cercanos o iguales a cero. Estos datos son generalmente números enteros provenientes de conteos de eventos aleatorios que siguen la distribución de Poisson. Por ejemplo, el número de colonias de bacterias contadas o leídas mediante el microscopio, el número de impurezas en muestras de leche, el número de plantas, aves o insectos de una especie en un sitio de muestreo; también datos porcentuales basados en conteos y con igual denominador, generalmente pequeños —entre 0 y 20%— o grandes —entre 80 y 100%— pero restados de 100 antes de transformarse. Por otra parte, la asimetría en los datos también puede corregirse con la transformación raíz cuadrada.

Normalmente es difícil interpretar los resultados del ANOVA sobre la escala transformada. Hinkelmann y Kempthorne (1994) dicen que la mejor estimación de las medias de los tratamientos sobre la escala original, se obtiene aplicando la transformación inversa a las medias calculadas con los datos transformados. De igual forma, los intervalos de confianza deben calcularse con la variable transformada y luego volver los límites a la escala original para su interpretación. Es necesario aclarar que las medias u otros valores reconvertidos a la escala original mediante la transformación inversa, son diferentes de los correspondientes valores originales y,

muchas veces, estos valores reconvertidos son estimaciones sesgadas. Así mismo, no es apropiado aplicar la transformación inversa a las desviaciones estándar ni a las varianzas, pues su interpretación es compleja.

Para reconvertir una media a la escala original en la transformación raíz cuadrada, esta debe elevarse al cuadrado; el resultado será un valor un poco inferior a la media aritmética de los datos originales. Si se usa $\sqrt{Y + 0.5}$, debe restarse 0.5 al valor reconvertido.

Ejemplo 5.3 El número de gusanos contados por parcela después de haber fumigado el suelo, el año inmediatamente anterior, con las sustancias K, M, N, O y P, es el que se muestra en la tabla 5.4 (adaptado de Snedecor y Cochran, 1984).

Tabla 5.4 Número de gusanos contados por parcela para cada uno de los tratamientos aplicados

	K	*M*	*N*	*O*	*P*
	4	6	17	4	3
	0	9	4	21	2
	1	8	5	9	2
	1	4	4	6	6
	0	4	5	8	4
Medias	1.2	6.2	7.0	9.6	3.4

Los supuestos del ANOVA para estos datos no se satisfacen a cabalidad. La prueba de Bartlett tiene nivel de significación $p = 0.016$ y los residuos presentan gran alejamiento de la normalidad. Si se transforman los datos a la escala $\sqrt{Y + 0.5}$, la homogeneidad de las varianzas resulta aceptable ($p = 0.38$) al igual que la normalidad. El ANOVA para los datos transformados es el de la tabla 5.5.

Tabla 5.5 Tabla del ANOVA para los datos transformados del ejemplo 5.3

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Tratamientos	10.465	4	2.616	2.49	0.0038
Error	9.528	20	0.476		
Total	19.993	24			

Las medias de los datos transformados y reconvertidas a la escala original son:

	K	M	N	O	P
Medias de los datos transformados:	1.197	2.558	2.623	3.061	1.941
Medias reconvertidas:	0.93	6.04	6.38	8.87	3.27

Estas últimas medias son inferiores a las medias no transformadas. Ahora bien, una corrección por sesgo consiste en sumar la *MCE* a cada media reconvertida; las medias así corregidas son: 1.4, 6.5, 6.9, 9.4 y 3.8, respectivamente.

Debe recordarse que los intervalos de confianza para las medias así como las comparaciones múltiples (véase capítulo 6), son válidos en la escala transformada y no en la original.

Transformación a la escala logarítmica

Esta transformación es apropiada cuando el coeficiente de variación $s_i / \bar{y}_{.j}$ es aproximadamente constante entre los grupos. Cuando hay ceros o datos pequeños, es preferible la transformación log(Y + 1) en vez de log(Y). Puede usarse cualquier base para los logaritmos, pero las bases natural y decimal son las más comunes.

La transformación logarítmica vuelve aditivos los efectos que son multiplicativos en la escala original. Esta transformación es aplicable a números enteros que varían en un rango amplio, y permite corregir la asimetría positiva de los datos. También es aplicable a datos experimentales relacionados con el crecimiento, como la variación en el tiempo del peso de los animales causada por dietas especiales.

La transformación inversa de la media aritmética, en este caso, su antilogaritmo, es la media geométrica de los datos originales. Por otro lado, el antilogaritmo de la diferencia de dos medias aritméticas en los datos transformados, es el cociente de la medias geométricas en la escala original.

Ejemplo 5.4 En un estudio de diferentes empaques para la carne, Gómez y González (1991) realizaron una prueba microbiológica y contabilizaron el número de mesófilos en las diferentes muestras analizadas. Los resultados se presentan en la tabla 5.6.

Nótese que estos datos no satisfacen los supuestos del ANOVA; en efecto, la prueba de homogeneidad de varianzas de Bartlett es significativa con $p = 1.37 \times 10^{-8}$ y los residuos no presentan normalidad, como se aprecia en la figura 5.10. Aunque el coeficiente de variación no es constante para los diferentes empaques, de todos modos tiene menos variabilidad que las respectivas varianzas, lo que sugiere ensayar una transformación logarítmica antes de realizar el ANOVA.

Se aplica la transformación ln(Y + 1) a los datos y se chequean nuevamente los supuestos. La prueba de Bartlett tiene significancia de $p = 0.017$ y la prueba de Kolmogorov-Smirnov, para la normalidad de los residuos, es no significativa con $p = 0.12$. La gráfica en el papel probabilístico Normal (figura 5.11), muestra menos alejamiento de la normalidad que la gráfica de la figura 5.10.

Tabla 5.6 Número de mesófilos en cada una de las diferentes envolturas

	Shopack	*Icopor*	*Biopack*	*Cryovac*	*Bolsa pp*	*Pel. recu.*
	280	10	80	10	10	30
	30	20	90	10	20	40
	0	20	10	10	40	40
	10	20	90	20	10	40
	50	30	60	0	10	10
Medias	74	20	66	10	18	32
DE	116.7	7.1	33.6	7.1	13.0	13.0
CV (%)	158	35	51	71	72	41
Medias de los ln	3.080	2.993	3.985	2.048	2.790	3.394
Medias geométricas	20.8	18.9	52.8	6.8	15.3	28.8

Efectuado el ANOVA que se muestra en la tabla 5.7, se concluye que no hay diferencias en las envolturas con respecto al número de mesófilos. En la tabla 5.6 aparecen las medias en la

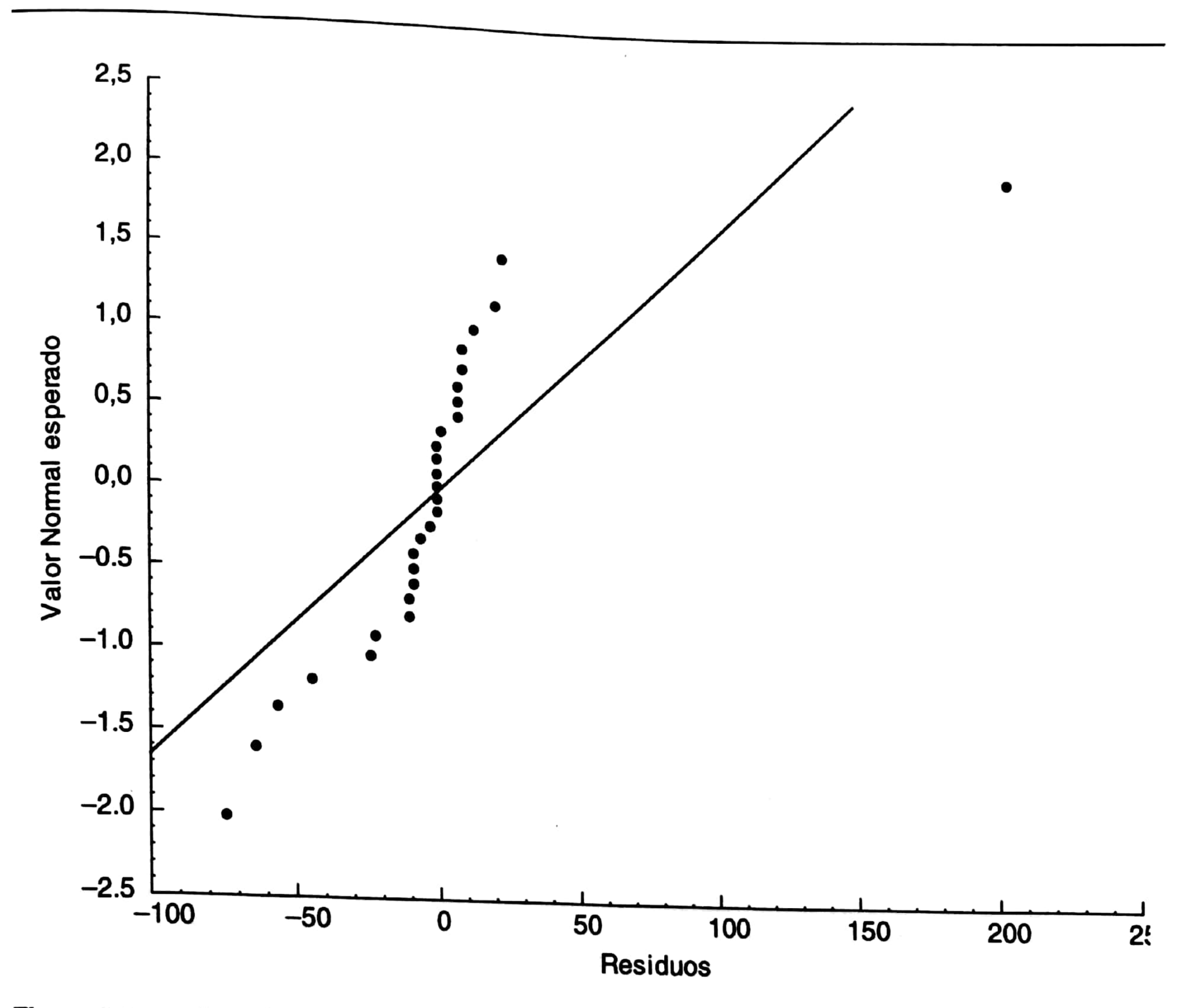


Figura 5.10 Gráfico de residuos para el ejemplo 5.4

escala logarítmica y también las medias geométricas en la escala original. Se observa que estas medias son inferiores a las medias aritméticas de los datos originales. Por ejemplo, la media para Shopack es $e^{3.080} - 1 = 20.8$. El dato 280 es un *outlier*, pero su exclusión no altera la conclusión anterior, esto es, la comparación de envolturas sigue siendo no significativa ($p = 0.059$).

Tabla 5.7 Tabla del ANOVA para los logaritmos naturales de los datos

Fuente de variación	*SC*	*gl*	*MC*	*F*	*P*
Entre envolturas	10.346	5	2.07	1.69	0.18 ns
Error	29.394	24	1.22		
Total	39.740	29			

Transformación a la escala angular

La transformación: $\theta = \text{arsen}\left(\sqrt{p}\right)$ conocida como transformación angular —θ representa un ángulo—, fue ideada para estudiar proporciones p de tipo binomial, como la proporción de semillas que germinan por parcela sembrada, o la proporción de

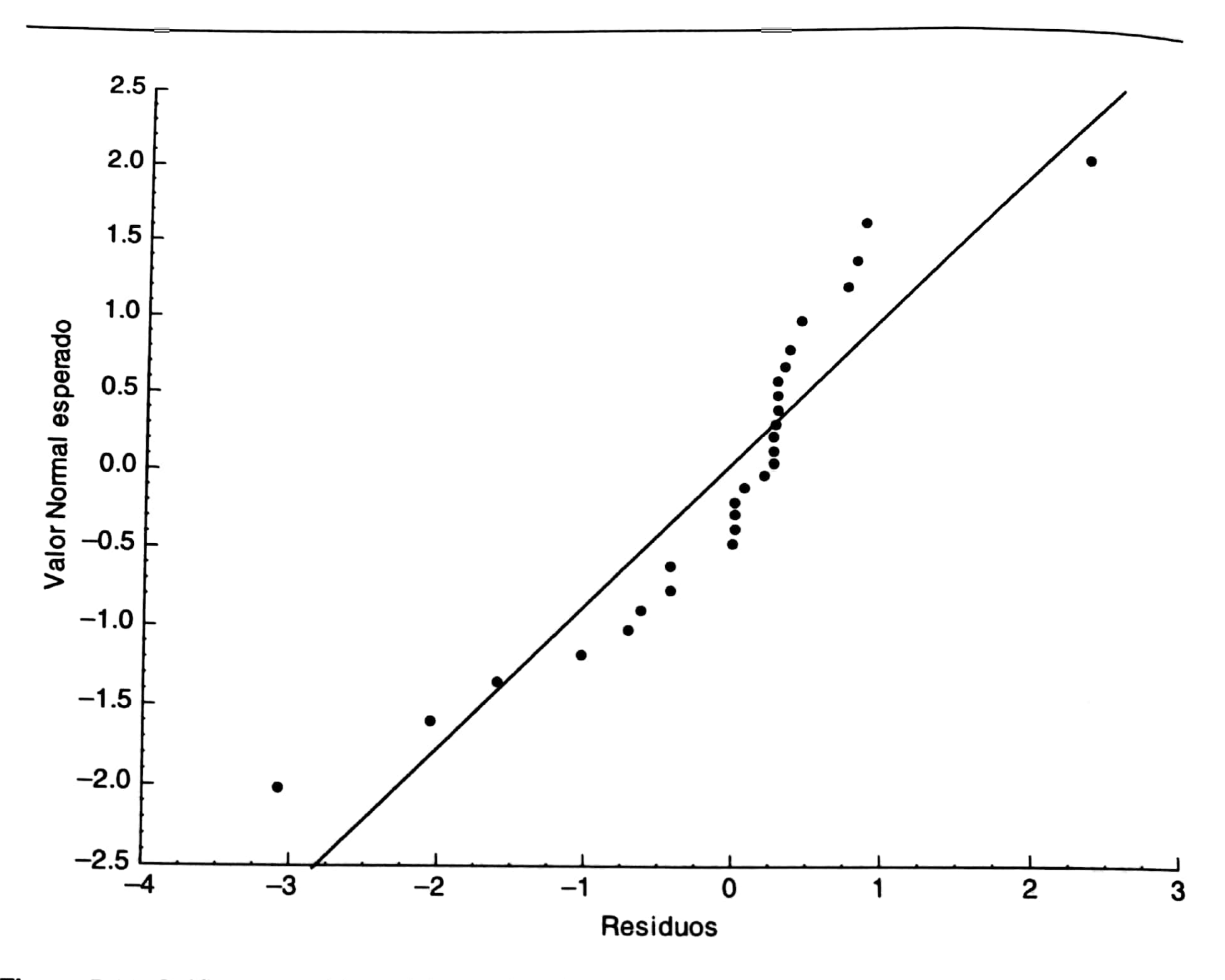


Figura 5.11 Gráfico de residuos de los ln(Y + 1) para el ejemplo 5.4

muertes de una especie de insectos sometidos a un agente químico. La varianza de los datos se estabiliza alrededor de 821 / n si θ está en grados o 0.25 / n si θ está en radianes; n es el denominador común de todas las fracciones. Si los porcentajes se calculan con denominadores diferentes, se requiere introducir ponderaciones en los datos.

Esta transformación es útil cuando los porcentajes cubren un amplio rango de valores. Si hay valores extremos, Bartlett (1947) recomienda reemplazar el cero por 1/4 y el n por n – 1/4. La transformación no es necesaria cuando los porcentajes se encuentran entre el 30 y el 70%, ya que en este intervalo no se produce el efecto esperado. De otro lado, existen tablas que permiten hacer la transformación directamente; Schefler (1981) por ejemplo, trae tablas con valores en grados. Sin embargo, hoy día las calculadoras con funciones trigonométricas permiten agilizar este trabajo.

Transformación a otras escalas

Se ha sugerido un número grande de otras transformaciones, tanto para homogeneizar varianzas como para normalizar los residuos o para obtener aditividad. Por

ejemplo, cuando la desviación estándar es proporcional al cuadrado de la media, la transformación $1/Y$ es apropiada para estabilizar las varianzas. Si los datos originales son conteos, se recomienda la transformación $1/(Y + 1)$. La transformación a la escala cuadrática se recomienda cuando hay asimetría negativa o cuando la desviación estándar y la media varían inversamente. Adicionalmente, hay otras dos transformaciones muy útiles, sobre todo en los ensayos biológicos, que se aplican a las proporciones: son las transformaciones *probit* y *logit*. Aunque estas no estabilizan varianzas tienen ventajas que las hacen interesantes, Finney (1978).

La posibilidad de aplicar una transformación a los datos amplía enormemente la clase de problemas que pueden resolverse por los métodos paramétricos. En esta dirección, Box y Cox (1964) han propuesto una transformación general en función de Y^{λ}, la cual tiene en cuenta los requisitos del ANOVA en la estimación del parámetro λ. En particular, $\lambda = -1$ es la transformación recíproca; $\lambda = 0$ se interpreta como la transformación logarítmica, pues Y^{λ} se comporta como $\log Y$ cuando λ es pequeña; y $\lambda = 0.5$ es la transformación raíz cuadrada. Zar (1999) proporciona métodos detallados para determinar la mejor transformación aplicable a datos heteroscedásticos.

Se dijo anteriormente que el ANOVA es robusto y que deben transformarse los datos únicamente cuando las desviaciones de los supuestos sean bastante severas, porque en este caso, el ANOVA puede conducir a conclusiones espurias. Sin embargo, no siempre es fácil lograr una transformación adecuada, y en ocasiones deberán ensayarse varias transformaciones antes de decidirse por una de ellas. Pero como dice Gill (1978), es importante considerar posibles correcciones a los supuestos que no se cumplen; pero también es importante reconocer que los supuestos más peligrosos pueden no ser los de los métodos estadísticos, sino los implícitos en la mente del científico y que lo conducen, casi ciegamente, a las inferencias subconscientemente deseadas o esperadas.

Finalmente, es conveniente explorar e identificar el o los posibles tratamientos causantes del problema. Muchas veces sus valores medios son notablemente diferentes de los demás y, en este caso, una corrección satisfactoria consiste en omitir estos tratamientos del análisis principal y realizar un ANOVA en la forma usual con los tratamientos restantes. Es claro que las conclusiones sobre los tratamientos excluidos pueden obtenerse de inmediato.

5.4 Prueba de rangos de Kruskall y Wallis

Los métodos estadísticos clásicos están basados en supuestos específicos sobre la distribución de la población objeto de la inferencia. Cuando estos supuestos no se cumplen, pueden resultar viciados tanto los contrastes de hipótesis como las estimaciones. A veces no se encuentra una transformación que ajuste los datos al modelo, por lo que se debe recurrir a otras técnicas de análisis. Entre las alternativas disponibles se tienen los métodos no paramétricos y los de libre distribución.

Para un diseño completamente aleatorizado, Kruskal y Wallis (1952) propusieron una prueba similar a la del análisis de varianza de una vía. Esta prueba asume el mismo modelo matemático del ANOVA, pero solo requiere que los errores sean

independientes —diseño aleatorizado— y que los datos provengan de una población continua, Hollander y Wolfe (1999).

Las hipótesis no se plantean en términos de parámetros sino de distribuciones, así:

H_0: las k muestras aleatorias fueron extraídas de k poblaciones idénticas.

H_1: las k poblaciones tienen distribuciones diferentes.

Si se concluye H_0, se acepta que los valores promedio poblacionales son iguales; en efecto, si las distribuciones son iguales, sus medias y varianzas también lo son. El procedimiento de prueba consta de los siguientes pasos:

1. Se asignan rangos, de menor a mayor, a los datos tomados en conjunto como una sola muestra. Si hay empates, la paridad se resuelve asignando el rango promedio a los datos iguales.
2. Se suman los rangos en cada grupo o muestra, siendo R_j la suma de los rangos del tratamiento j con r_j repeticiones.
3. Se calcula la estadística

$$H = \left[12/n(n+1)\right]\sum_j \frac{R_j^2}{r_j} - 3(n+1), \text{ donde } n = \sum_j r_j.$$

4. La regla de decisión consiste en rechazar H_0 si $H > \chi^2_{\alpha,k-1}$.

Existen tablas que dan la probabilidad exacta para valores de H cuando $k = 3$ y $r_j \leq 5$. Si $r_j > 5$, la distribución de la estadística H tiende a la distribución *ji*-cuadrado con $k - 1$ grados de libertad.

5. Si el número de empates es grande, se recomienda usar un factor de corrección que aumente el valor de H, el cual puede tener alguna influencia cuando se acepta H_0.

En algunas investigaciones, la variable de respuesta solo alcanza a medirse en la escala ordinal. Pero los datos ordinales no suelen analizarse con métodos paramétricos, así que para estos problemas el método de Kruskall y Wallis parece ser el más adecuado. Inicialmente se estableció que la población debe ser continua, pero este requisito se fijó con miras a evitar los empates. Sin embargo, aunque la variable sea continua, la escala de medición puede ser solamente ordinal.

A pesar de que el método de Kruskall y Wallis es más rápido y fácil de aplicar, surge la pregunta de por qué no se usa con mayor frecuencia. En primer lugar, para $r_j \leq 5$, la estadística H tiene distribución apenas aproximada a la *ji*-cuadrado y para $k > 3$ no existen tablas que den las probabilidades directamente. Por otra parte, si el ANOVA es aplicable, la potencia de la prueba F es generalmente mayor que la correspondiente al método de distribución libre. Además, algunos estadísticos arguyen que la prueba de Kruskall y Wallis no es buena alternativa para la F cuando las varianzas son demasiado heterogéneas.

Ejemplo 5.5 Gómez y González (1991) investigaron la pérdida de peso de varios tipos de carne después de permanecer cinco días refrigerada y empacada en biopak. Para cada tipo de carne hicieron cinco repeticiones y calcularon la pérdida de peso en porcentaje con respecto al peso inicial. Los resultados se dan en la tabla 5.8

Tabla 5.8 Pérdida de peso en porcentaje de diferentes tipos de carne después de cinco días de almacenamiento

Milanesa de cerdo	*Posta*	*Milanesa de res*	*Tocino*	*Costilla*
2.15 (15)	2.17 (16)	6.59 (23)	0.89 (3)	1.11 (8.5)
4.84 (20)	2.08 (14)	7,89 (25)	1.10 (7)	1.19 (10)
2.90 (17)	1.54 (13)	4.48 (19)	0.91 (4)	1.11 (8.5)
3.66 (18)	1.39 (12)	7.32 (24)	1.02 (5)	1.09 (6)
5.88 (21)	0.70 (1)	6.41 (22)	0.82 (2)	1.20 (11)
R_j (91)	(56)	(113)	(21)	(44)

El ANOVA no es aplicable a estos datos porque, aunque el supuesto de normalidad sea aceptable, el de homogeneidad de varianzas está lejos de cumplirse —Bartlett, $p = 8.4 \times 10^{-7}$ y Kolmogorov-Smirnov, $p = 0.13$—. Además, ninguna de las transformaciones más comunes mejora sensiblemente los requisitos iniciales. Se aplica, entonces, el método de Kruskall y Wallis para comparar la pérdida de peso en las diferentes carnes.

Cada dato tiene al lado el rango correspondiente (véase tabla 5.8); así, el 0.70 tiene el rango 1, por ser el valor mínimo entre todos. El dato 1.11 está repetido dos veces y recibiría los rangos 8 y 9; sin embargo, como están empatados, se les asigna el rango promedio, es decir, 8.5.

El valor de la estadística H se calcula con base en las sumas de los rangos. Para $n = 25$ y $r_j = 5$, se tiene:

$$H = [12/(25 \times 26)](1/5)[(91)^2 + (56)^2 + (113)^2 + (21)^2 + (44)^2] - 3 \times 26 = 20.08.$$

El valor tabulado $\chi^2_{0.05,4} = 9.44 < H$, o sea que existen diferencias en la pérdida de peso para los diferentes tipos de carne ($p = 0.0005$). Estas diferencias se analizarán en la sección 6.12.

5.5 Medias α-recortadas en análisis de varianza

En algunos problemas de investigación, los datos experimentales pueden presentar simetría y homogeneidad de varianzas en los diferentes grupos del ANOVA. Sin embargo, la presencia de muchos datos extremos origina largas colas en sus distribuciones, lo cual da como resultado muy poca sensibilidad en las pruebas. Una forma de aumentar la sensibilidad sin perder validez, consiste en aplicar la inferencia a las medias recortadas, esto es, suprimir o recortar un porcentaje α de los datos extremos en cada tratamiento o muestra.

Una media recortada el 10%, por ejemplo, se obtiene eliminando el 10% de las observaciones en cada extremo del conjunto ordenado de los datos, y calculando la media aritmética de los datos restantes. La metodología de eliminar datos extremos ha sido utilizada durante más de dos siglos, pero solo a finales del siglo pasado fue sometida a escrutinio estadístico. Se encontró que era robusta, tanto a la validez como a la sensibilidad de las inferencias, en presencia de distribuciones de colas largas, Tukey y McLaughlin (1963).

Los procedimientos de estimación basados en $\bar{y}$ y s^2 son robustos en cuanto a validez, pero no en cuanto a sensibilidad. El problema se debe a la fuerte influencia

que ejercen los datos extremos sobre la varianza. Cuando se da este caso, los datos extremos pueden eliminarse utilizando el método de medias recortadas, que se explica a continuación.

Ejemplo 5.6 Los siguientes datos fueron mediciones hechas, en un estudio sobre control de contaminación ambiental, de la cantidad de hidrocarburos, HC, emitidos por automóviles de diferentes años de fabricación. Los datos se muestran en la tabla 5.9 (adaptado de Koopmans, 1987).

La gráfica de las cajas esquemáticas (véase figura 5.12) indica la existencia de valores extremos y de heterogeneidad en las varianzas. Con el fin de disminuir el efecto de los datos extremos, se aplica el procedimiento de las medias recortadas a estos datos y se recorta el 10% en cada extremo de los grupos o muestras.

La fracción que se debe recortar en un problema dado no es fija. Se recomienda entre el 5 y el 25%, dependiendo del grado de contaminación —número relativo de valores extremos— que presentan las muestras. La idea es recortar hasta eliminar todos los valores extremos; este objetivo se logra, con frecuencia, haciendo recortes del 10%.

Considérese la última muestra, años 1972-1974. Su tamaño es $n = 14$ y el 10% de recorte es 1.4, que se aproxima al mayor entero contenido en el número, o sea, se recorta un dato en cada extremo. La muestra ordenada es la siguiente:

1880, 400, 360, 223, 200, 175, 160, 140, 85, 70, 60, 20, 20, 20.

Se suprimen los valores subrayados en los extremos. El nuevo tamaño muestral es $h = 12$, y la media recortada es:

$$\bar{y}_T = \sum y_i / h = 1913/12 = 159.4.$$

El error estándar de la media recortada se calcula a partir de la muestra winsorizada, la cual se obtiene reemplazando los datos recortados por el dato más extremo que queda después de hacer el recorte. La muestra winsorizada es:

400, 400, 360, 223,..., 70, 60, 20, 20, 20.

Tabla 5.9 Emisión de hidrocarburos (en ppm) por automóviles de diferentes años de fabricación

Antes de 1963	*1963-1967*	*1968-1969*	*1970-1971*	*1972-1974*
1293	620	1088	141	140
541	940	388	359	160
1058	350	111	247	20
411	700	558	940	20
570	1150	294	494	223
2351	2000	211	306	60
800	823	460	100	20
630	1058	470	300	360
905	423	353	190	70
347	270	71	140	400
	900	117	880	1880
	405	2999	223	200
	780		188	175
			241	85

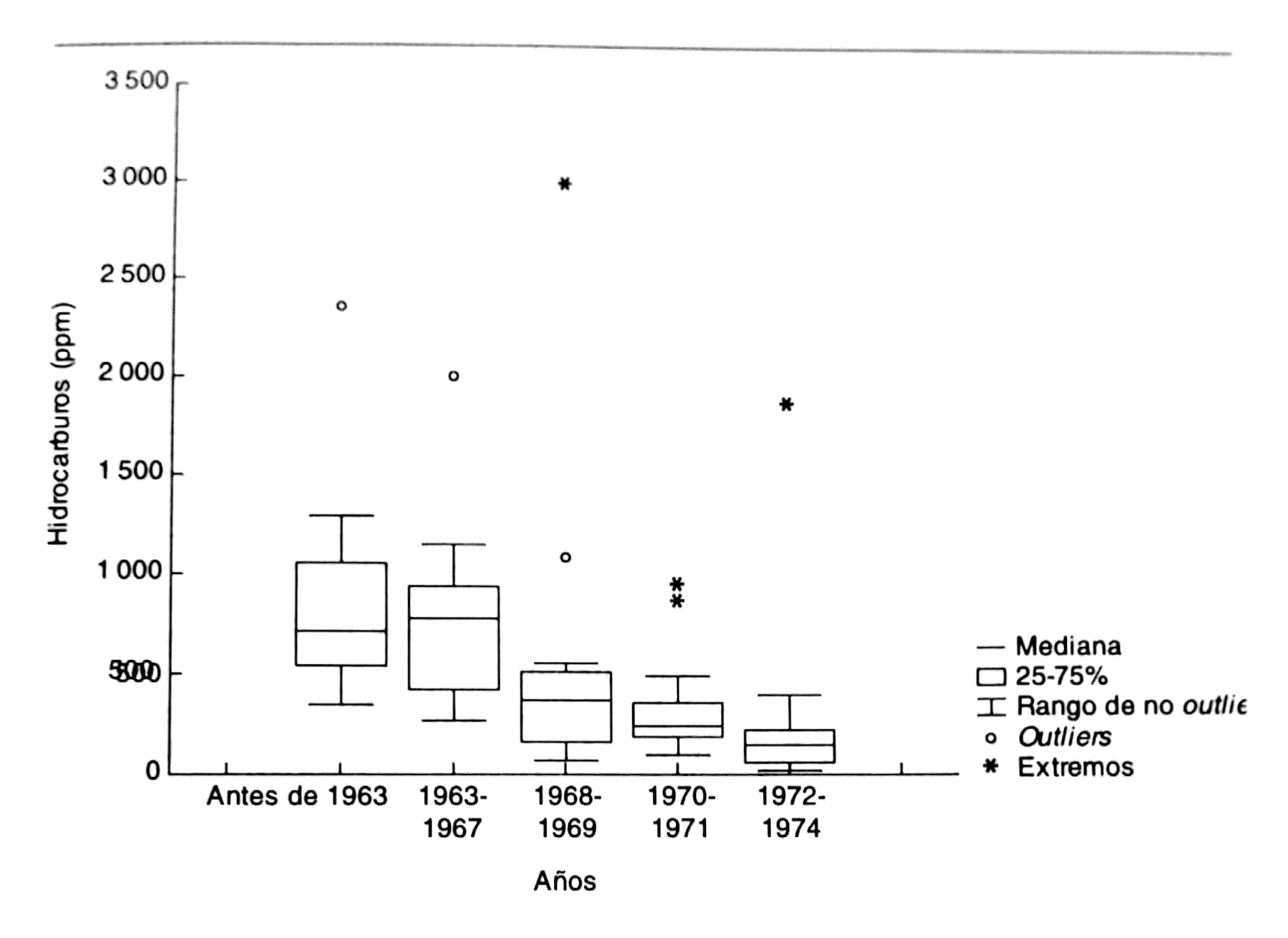


Figura 5.12 Cajas esquemáticas para el ejemplo 5.6

La desviación estándar de esta muestra es s = 136.80 y la desviación estándar recortada es:

$$s_T = s\sqrt{(n-1)/(h-1)}, \text{ o sea, } s_T = 136.80\sqrt{13/11} = 148.7.$$

Las medias y desviaciones estándares recortadas el 10% para cada muestra se presentan en la tabla 5.10.

Realizando el ANOVA para las muestras recortadas, se encuentra que hay diferencias significativas ($p < 0.001$) entre los promedios de las emisiones de hidrocarburos de los vehículos según el año de fabricación, como se muestra en la tabla 5.11.

Ejercicios

5.1 Con respecto al ejercicio 4.5, sobre concentración de calcio en hojas:

a. Calcular los residuos e_{ij}.
b. Construir un diagrama lineal de puntos para estos residuos y explorar su normalidad.
c. Graficar los residuos contra orden, en el tiempo, y analizar independencia.

Tabla 5.10 Medias y desviaciones estándares recortadas el 10%

	Antes de 1963	*1963-1967*	*1968-1969*	*1970-1971*	*1972-1974*
h	8	11	10	12	12
$\bar{y}_T$	776.0	740.8	399.0	309.1	159.4
s_T	380.7	323.2	375.5	270.9	148.7

Tabla 5.11 Tabla del ANOVA para las muestras recortadas del ejemplo 5.6

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Años de fabricación	3057832	4	764458	13.6	<0.001
Error	2696131	48	56169		
Total	5753963	52			

d. Graficar los residuos contra los valores estimados y analizar homogeneidad de las varianzas de los tratamientos.
e. Aplicar una prueba de normalidad a estos residuos. Dar el valor de p.
f. Aplicar la prueba de Bartlett para chequear la homogeneidad de las varianzas. Dar el valor de p.

5.2 Chequear los supuestos de homogeneidad de varianzas, independencia de residuos y normalidad de residuos para los datos de los ejercicios 4.4, 4.6, 4.8 y 4.10. Para estos chequeos se recomienda usar *software* estadístico apropiado.

5.3 (Opcional). Antes de analizar una tabla del ANOVA deben chequearse los supuestos del modelo matemático. Esto se hizo con los ejercicios del capítulo anterior. No obstante, al resolverlos por primera vez, es importante comprobar que los supuestos sí se cumplen.

5.4 En el ejercicio 4.9 sobre células productoras de insulina:

a. Chequear los supuestos del ANOVA.
b. Realizar el ANOVA para medias recortadas. Mirar las cajas esquemáticas según el ejercicio 4.9*b* para decidir sobre la fracción que se debe recortar.

5.5 Moreno y otros (1992) investigaron la relación entre la periodontitis juvenil localizada, PJL, y el comportamiento in vitro de los polimorfonucleares neutrófilos, PMN. Se analizaron 21 pacientes con PJL clasificados en tres grupos: incipiente, moderada y avanzada. Se estudió la actividad metabólica de los PMN estimulados con zimosan opsonizado mediante una prueba de quimioluminiscencia dependiente del luminol, QLDL. Se registraron siete ciclos de lectura en intervalos de diez minutos. Los resultados siguientes, donde la energía lumínica está medida en cpm, corresponden a los registros diez minutos después de iniciado el proceso.

Controles	*PJL incipiente*	*PJL moderada*	*PJL avanzada*
814 170	910 300	606 100	700 040
936 650	664 880	475 750	158 340
765 130	222 420	904 670	365 490
434 260	641 980	1 363 130	554 320
702 650	1 221 780	425 880	436 940
464 760		2 304 780	1 815 980
651 220		232 160	1 042 280
1 113 470		629 410	309 300
883 240			
591 770			
735 610			
709 070			
719 100			
1 067 640			
1 066 560			

a. Especificar el factor de estudio y sus niveles, y la variable de respuesta.
b. Especificar el diseño y el modelo apropiado para este estudio.
c. Establecer las hipótesis del problema.
d. Chequear los supuestos del modelo y, si es necesario, aplicar la transformación adecuada. ¿Podría pensarse en medias α-recortadas?
e. Efectuar el ANOVA y obtener conclusiones

5.6 En el estudio realizado por Gómez y González (1991) con diferentes tipos de carnes, se midió el porcentaje de pérdida de peso de la costilla de cerdo después de cinco días de empacada. A continuación se muestran los resultados.

Icopor	*Biopak*	*Cry-o-vac*	*Shopak*
1.10	1.11	1.75	1.01
0.00	1.19	4.35	0.00
0.00	1.11	1.43	0.64
0.00	1.09	2.44	0.78
0.00	1.20	2.13	2.06

a. Comprobar que las varianzas son heterogéneas.
b. Comprobar que las transformaciones conocidas no destruyen la heterogeneidad.
c. Resolver el problema usando el método de Kruskall y Wallis.

5.7 Se compararon tres métodos para reducir la flora bacteriana de la piel. Los conteos de las bacterias se efectuaron en el pie derecho de personas,

antes y después del tratamiento. La variable de respuesta fue el porcentaje de disminución de las bacterias (Daniel, 2004).

Método	*Repeticiones*								
Agitación por remolino	91	87	88	84	86	80	92	81	93
Duchado	18	22	20	29	25	16	15	26	19
Baño de pies	6	6	8	9	13	10	12	5	9

a. Observar los datos y sacar conclusiones.
b. Aplicar el método de Kruskall y Wallis para corroborar lo observado en *a*.

5.8 Se estudiaron 22 jóvenes asmáticos para establecer los efectos a corto plazo del dióxido de azufre (SO_2). Los datos obtenidos representan la reactivación bronquial (en cm H_2O/s) al SO_2. Los jóvenes fueron agrupados de acuerdo con una función pulmonar (FEV_1/FVC) en tres niveles (Rosner, 1999):

Grupo A: ≤74%:	20.8	4.1	30.0	24.7	13.8						
Grupo B: 75-84%:	7.5	7.5	11.9	4.5	3.1	8.0	4.7	28.1	10.3	10.0	5.1	2.2
Grupo C: ≥85%:	9.2	2.0	2.5	6.1	7.5						

a. Analizar las diferencias entre los grupos usando el método de Kruskall y Wallis.
b. Discutir el uso del ANOVA de una vía para resolver este problema en vez del método no paramétrico.

6

Comparaciones múltiples

6.1 Contrastes entre medias de tratamientos

En los problemas de investigación donde se plantea el estudio de tres o más tratamientos, es usual que el investigador formule un conjunto de comparaciones y contrastes para verificarlos estadísticamente. El ANOVA puede considerarse como un primer paso en la metodología del análisis estadístico de estas comparaciones.

Aquí se supondrá que se tiene un solo factor de análisis con k niveles o tratamientos seleccionados previamente, de modo que sea aplicable el modelo de efectos fijos. Para diseños factoriales, el análisis es similar si se supone que el factor en estudio tiene niveles fijos.

Cuando la prueba F del ANOVA es significativa, se concluye que hay diferencias entre las medias de los tratamientos; sin embargo, el ANOVA por sí mismo no indica cuáles son esas diferencias. Las hipótesis H_0: $= \mu_1 = \mu_2 = \cdots = \mu_k$ y H_1: no todas las μ_j son iguales, no proporcionan respuestas específicas a cualquier comparación entre medias, ya que la hipótesis alternativa es una hipótesis compuesta.

El problema de las estimaciones y pruebas de hipótesis sobre comparaciones entre tratamientos, ha llamado la atención de muchos estadísticos y, actualmente, se dispone de una gran variedad de procedimientos para enfrentar su solución, los cuales se conocen como métodos de comparaciones múltiples.

Las comparaciones múltiples entre tratamientos se clasifican en comparaciones diseñadas y en comparaciones sugeridas por los datos o no diseñadas. Las primeras, conocidas también como comparaciones a priori, son las que el investigador formula y plantea cuando está diseñando el experimento, son comparaciones interesantes y con sentido en la investigación; mientras que las segundas, conocidas como comparaciones a posteriori, surgen después de conocer y escudriñar los datos; por ejemplo, la comparación entre los tratamientos que presentaron resultados extremos, esto es, el promedio mayor comparado con el menor.

Lo ideal es que el investigador conozca a fondo el fenómeno y formule, durante el diseño, aquellas comparaciones o contrastes que permitan responder a objetivos específicos. En la práctica no siempre se tiene claridad sobre qué tipo de contraste

es interesante, y muchas veces, para ello, es necesario esperar hasta conocer los resultados experimentales. El más común y el más analizado es la comparación de medias por pares.

Supóngase, por ejemplo, que un experimento se diseñó para comparar el efecto nutricional de varias raciones de alimento en la producción de leche de vacas lecheras. Los tratamientos seleccionados, esto es, los alimentos, fueron: 1) el estándar con el 9% de proteína, 2) el estándar con el 13% de proteína, 3) urea con el 9% de proteína, 4) urea con el 13% de proteína y 5) amonio con el 13% dé proteína. Algunas preguntas que el investigador puede formularse en este experimento son:

1. ¿Se diferencia el promedio entre la urea y el amonio del promedio del alimento estándar, como fuente de nitrógeno en raciones con un 13% de proteína?
2. ¿Son la urea y el amonio diferentes entre sí al aplicarse con el 13% de proteína?
3. ¿Es la urea equivalente al alimento convencional en raciones con el 9% de proteína?
4. ¿Es igual la producción de leche con el alimento estándar que con el amonio?
5. Si se usa urea, ¿sería igual la producción de leche con raciones al 9% que con raciones al 13% de proteína?

Cada uno de los k tratamientos en estudio origina una subpoblación que tendría el valor promedio μ_j, ($j = 1, 2, \ldots, k$). Algunas comparaciones entre estos valores promedio reciben el nombre de contrastes.

Definición: un contraste es una comparación entre las medias de los tratamientos expresada como una combinación lineal de la forma:

$$Q = \sum_{j=1}^{k} c_j \mu_j,$$

donde los coeficientes c_j deben satisfacer la condición $\sum_j c_j = 0$. Se supone que cada tratamiento está repetido r veces, es decir, que se tiene un diseño balanceado.

6.2 Contrastes ortogonales

Dos contrastes $Q_1 = \sum_j c_j \mu_j$ y $Q_2 = \sum_j d_j \mu_j$ son ortogonales o independientes solo si $\sum_j c_j d_j = 0$; en caso contrario son no ortogonales (están correlacionados).

En el ejemplo sobre nutrición animal, cada una de las preguntas formuladas puede expresarse como un contraste entre medias. Sean $\mu_1, \mu_2, \ldots, \mu_5$ las medias de los tratamientos en el orden citado inicialmente; entonces, la primera pregunta equivale a comparar μ_2 —alimento convencional con 13% de proteína— con el promedio entre μ_4 y μ_5 —urea y amonio, cada uno con 13% de proteína. La comparación se establece mediante la diferencia:

$\mu_2 - (\mu_4 + \mu_5) / 2$, o también como $2\mu_2 - \mu_4 - \mu_5$.

Esta relación entre medias es un contraste puesto que los coeficientes c_j son:

$$c_1 = 0, c_2 = 2, c_3 = 0, c_4 = -1, c_5 = -1 \text{ y}$$
$$\sum_j c_j = 0 + 2 + 0 - 1 - 1 = 0 .$$

Para responder la primera pregunta es necesario plantearla como una hipótesis estadística en la forma H_0: $\mu_2 = (\mu_4 + \mu_5) / 2$ contra H_1: $\mu_2 \neq (\mu_4 + \mu_5) / 2$. La hipótesis nula puede formularse también como H_0: $2\mu_2 - \mu_4 - \mu_5 = 0$, o simplemente como H_0: $Q_1 = 0$, donde Q_1 es el contraste $0\mu_1 + 2\mu_2 + 0\mu_3 - 1\mu_4 - 1\mu_5$. Las cuatro preguntas restantes pueden expresarse en forma similar; es decir, como cuatro contrastes Q_i, y cada una se responde probando las hipótesis

$$H_0\text{: } Q_i = 0 \text{ contra } H_1\text{: } Q_i \neq 0 \text{ para } i = 1, 2, \dots, m.$$

La segunda pregunta compara a μ_4 con μ_5, de modo que el contraste es de la forma $Q_2 = \sum_j d_j\mu_j = 0\mu_1 + 0\mu_2 + 0\mu_3 + 1\mu_4 - 1\mu_5 = \mu_4 - \mu_5$. Los coeficientes son: $d_1 = 0, d_2 = 0, d_3 = 0, d_4 = 1, d_5 = -1$ y $\sum_j d_j = 0$. Los contrastes Q_1 y Q_2 son ortogonales porque

$$\sum_j c_j d_j = c_1 d_1 + c_2 d_2 + c_3 d_3 + c_4 d_4 + c_5 d_5$$
$$= (0 \times 0) + (2 \times 0) + (0 \times 0) + (-1)(1) + (-1)(-1) = 0.$$

Las cinco preguntas anteriores, expresadas como contrastes, están reunidas en la tabla 6.1. En la misma tabla aparecen los coeficientes c_{ji}, respectivos, donde i indica el orden del contraste.

Los tres primeros contrastes son ortogonales entre sí —lo son dos a dos—, el cuarto es ortogonal con el tercero y el quinto, mientras que este último lo es únicamente con el cuarto. Si los contrastes diseñados por el investigador son mutuamente ortogonales, el análisis estadístico es más sencillo: hay mayor sensibilidad o potencia en las pruebas y más precisión en las estimaciones que en el caso de contrastes correlacionados.

Una sugerencia para el investigador es que, en lo posible, formule contrastes ortogonales correspondientes a hipótesis específicas de interés. No obstante, hay dos restricciones fundamentales que no permiten seguir esta pauta: primera, hay contrastes que son de interés primario pero que resultan no ortogonales con los demás contrastes; segunda, el número de contrastes podría resultar mayor que el número máximo permitido para satisfacer la ortogonalidad.

Tabla 6.1 Contrastes definidos por las preguntas iniciales del investigador

Contrastes	c_{1i}	c_{2i}	c_{3i}	c_{4i}	c_{5i}
Q_1: estándar versus urea y amonio (13%)	0	+2	0	−1	−1
Q_2: urea versus amonio (13%)	0	0	0	−1	+1
Q_3: estándar versus urea (9%)	+1	0	−1	0	0
Q_4: estándar versus amonio (13%)	0	+1	0	0	−1
Q_5: urea (9%) versus urea (13%)	0	0	+1	−1	0

Cuando no se tiene ortogonalidad, existen otras alternativas de análisis a las que es necesario recurrir si se opta por no renunciar a algunas exigencias iniciales en el diseño. Gill (1978, vol. 1) sugiere los siguientes pasos antes de utilizar otras alternativas, que generalmente ofrecen menos sensibilidad en el análisis.

Cuando un contraste es no ortogonal con los demás, se debe reexaminar el objetivo que origina la pregunta para estudiar la posibilidad de definir un contraste un poco diferente, pero que resulte ortogonal a los otros y que responda satisfactoriamente al objetivo propuesto. Si esto no es posible, se debe calcular el grado de correlación con otros contrastes mediante la fórmula:

$$\rho_{ij} = \frac{\sum c_j d_j}{\sqrt{\sum c_j^2 \sum d_j^2}}.$$

Si las correlaciones entre los varios pares de contrastes son pequeñas, por ejemplo si $|\rho| \leq 0.20$, o si solamenate un par entre un conjunto mayor está correlacionado, puede escogerse entre ignorar la no ortogonalidad, esperando que las consecuencias sobre distorsiones de tipo probabilítico sean mínimas, o recurrir a procedimientos de análisis dirigidos a contrastes no ortogonales.

En el ejemplo en discusión, la correlación entre Q_1 y Q_4 es $\rho_{14} = 3\sqrt{6 \times 2} = 0.87$, entre Q_1 y Q_5 es $\rho_{15} = 0.29$ y entre los demás contrastes correlacionados es de -0.50, Una alternativa es formular Q_5, de otra manera, por ejemplo: todas las raciones con el 13% de proteína contra todas las raciones con el 9% de proteína. Este contraste es más general que Q_1 y Q_5, pero es ortogonal con cada uno de los tres primeros.

El número máximo de contrastes ortogonales posibles en un experimento es igual al número de tratamientos menos uno, esto es $k - 1$, Das y Giri (1986). Para el ejemplo, no más de cuatro contrastes podrían ser ortogonales; así, habría que eliminar uno si se desea la ortogonalidad. Este sería Q_4 porque *es el más correlacionado;* por supuesto, se deja el nuevo Q_5 como Q_4.

Otra alternativa, posiblemente menos drástica, puede ser la de anular Q_4 y dejar Q_5, ignorando la correlación de este con los otros contrastes para aprovechar la sensibilidad del procedimiento de contrastes ortogonales. Debe reconocerse, sin embargo, que la probabilidad de error de tipo I o el grado de confianza en la estimación de intervalos, pueden distorsionarse en forma desconocida. Este aspecto inferencial se discutirá con más detalle en la sección 6.4.

Das y Giri (1986) y Lindman (1992) demuestran que existen infinitos conjuntos de $k - 1$ contrastes mutuamente ortogonales. Uno de ellos, por ejemplo, puede formarse agregando progresivamente una media al contraste anterior, así:

1. $\mu_1 - \mu_2$.
2. $\mu_1 + \mu_2 - 2\mu_3$.
3. $\mu_1 + \mu_2 + \mu_3 - 3\mu_4$.
4. $\mu_1 + \mu_2 + \ldots$

...

Esto se hace hasta obtener los $k - 1$ contrastes que resultan ortogonales.

También existen tablas de contrastes ortogonales para varios valores de k, conocidas como tablas de polinomios ortogonales.

6.3 Análisis de contrastes ortogonales

Los contrastes entre medias de tratamientos pueden analizarse mediante la prueba t de Student, la prueba F o los intervalos de confianza. Un contraste, como ya se dijo, es una combinación lineal de medias μ_j —en general una combinación lineal de parámetros— que puede estimarse reemplazando las medias μ_j por sus estimadores $\bar{y}_{.j}$. Entonces, un estimador de Q es $\hat{Q} = \sum_j c_j \bar{y}_{.j}$, la cual es una variable aleatoria con valor esperado $E(\hat{Q}) = \sum_j c_j \mu_j = Q$ y varianza $V(\hat{Q}) = (\sigma^2/r)\sum_j c_j^2$, aquí σ^2 es la varianza común a todos los tratamientos estimada por la MCE del ANOVA y r es el número de repeticiones por tratamiento.

Para probar la hipótesis H_0: $Q = 0$, se calcula el cociente $F = SC(\hat{Q})/MCE$, donde $SC(\hat{Q}) = (r\hat{Q}^2)/\sum c_j^2$ con 1 gl. También puede calcularse $t = \hat{Q}/S(\hat{Q})$, donde $S(\hat{Q})$ es un estimador de la desviación estándar de $\hat{Q}$. Un intervalo de confianza para Q se construye como $\hat{Q} \pm (t_{\alpha/2,v})S(\hat{Q})$.

Para probar hipótesis sobre contrastes es preferible usar la prueba F, pero calculando el numerador como una combinación lineal de totales T_j en vez de medias $\bar{y}_{.j}$. En este caso $\hat{Q} = \sum_j c_j T_j$, $E(\hat{Q}) = r\sum_j c_j \mu_j$, $V(\hat{Q}) = r\sigma^2 \sum_j c_j^2$ y la suma de cuadrados de $\hat{Q}$ es $SC(\hat{Q}) = Q^2/r\sum_j c_j^2$.

Ejemplo 6.1 Se diseñó un experimento para comparar la longitud de las secciones de guisantes sembrados en un cultivo de tejidos, cuyo medio contenía auxina. El objetivo del experimento era contrastar los efectos de la adición de diferentes azúcares en el crecimiento, expresado por la longitud alcanzada. Se utilizaron cinco tratamientos: cuatro grupos experimentales y un control sin azúcar (véase tabla 6.2). Entre los grupos experimentales se ensayaron tres azúcares por aparte —glucosa, fructosa y sacarosa— y una mezcla de glucosa y fructosa. En la tabla 6.2, el control corresponde al primer tratamiento, el segundo tratamiento es glucosa al 2%, etc. Se hicieron diez repeticiones para cada tratamiento en un diseño completamente aleatorizado (adaptado de Sokal y Rohlf, 1995).

El experimento se diseñó para responder, específicamente, a las siguientes preguntas:

1. ¿Se afecta la longitud de las secciones de guisantes por la adición de azúcares?
2. ¿Habrá diferencias de longitud al adicionar azúcares puros y azúcares mezclados?
3. ¿Habrá diferencias entre agregar disacáridos —sacarosa— y monosacáridos —glucosa, fructosa, glucosa + fructosa—?
4. ¿Habrá diferencias entre glucosa y fructosa?

La tabla 6.3 resume las preguntas anteriores convertidas en contrastes y especifica los coeficientes de cada uno de ellos en el orden de los tratamientos.

Aunque la prueba F del ANOVA no es necesaria cuando en el diseño inicial se plantean varios contrastes, es interesante realizar el análisis de varianza completo como un primer paso, incluyendo los chequeos de los supuestos.

Para el análisis de contrastes se requiere la MCE, la cual puede leerse en la tabla 6.4. De esta se concluye que los tratamientos con azúcares tienen efecto negativo altamente significativo en el crecimiento de las secciones de guisantes. Sin duda, los azúcares producen efecto inhibitorio del crecimiento, pues reducen su longitud ($p < 0.001$).

Por otra parte, los contrastes definidos en la tabla 6.3 son mutuamente ortogonales excepto Q_2 y Q_3, cuya correlación es $\rho_{\hat{2}\hat{3}} = -0.33$. Esta correlación será ignorada en el análisis siguiente, pues es la única existente y además es relativamente pequeña.

Tabla 6.2 Longitudes de secciones de guisantes crecidas en cultivo de tejidos y en presencia de auxina. La longitud está dada en unidades de micrómetro ocular (x0.114 = mm)

Réplicas	*Control*	*Glucosa al 2%*	*Fructosa al 2%*	*Sacarosa al 2%*	*Glucosa al 1%, fructosa al 1%*
1	75	57	58	62	58
2	67	58	61	66	59
3	70	60	56	65	58
4	75	59	58	63	61
5	65	62	57	64	57
6	71	60	56	62	56
7	67	60	61	65	58
8	67	57	60	65	57
9	76	59	57	62	57
10	68	61	58	67	59
Totales T_j	701	593	582	641	580
Media $\bar{y}_{.j}$	70.1	59.3	58.2	64.1	58.0

Tabla 6.3 Contrastes correspondientes a las preguntas del experimento del ejemplo 6.1

Contrastes	c_{1i}	c_{2i}	c_{3i}	c_{4i}	c_{5i}
Q_1: $[\mu_1 - (1/4)(\mu_2 + \mu_3 + \mu_4 + \mu_5)]$	4	−1	−1	−1	−1
Q_2: $[(1/3)(\mu_2 + \mu_3 + \mu_4) - \mu_5)]$	0	1	1	1	−3
Q_3: $[(1/3)(\mu_2 + \mu_3 + \mu_5) - \mu_4)]$	0	1	1	−3	1
Q_4: $[\mu_2 - \mu_3]$	0	1	−1	0	0

Tabla 6.4 Tabla del ANOVA para los datos del ejemplo 6.1

Fuente de variación	*SC*	*gl*	*MC*	*F*
Entre tratamientos	1077.32	4	269.3	49.33
Error	245.50	45	5.46 (MCE)	
Total	1322.82	49		

Pruebas de hipótesis sobre contrastes

Las fórmulas dadas al principio de esta sección permiten estimar los contrastes definidos por medio de totales, y efectuar las pruebas de significancia correspondientes.

Este análisis se presenta, discriminado para cada uno de los contrastes de la tabla 6.3, en la tabla 6.5.

Si los contrastes fueran mutuamente ortogonales, las cuatro sumas de cuadrados de contrastes —cada una con 1 *gl*— totalizarían la suma de cuadrados entre tratamientos del ANOVA inicial, la cual está basada en 4 *gl*. La diferencia (44.38 = 1 121.70 – 1 077.32) se debe a la correlación entre Q_2 y Q_3.

Las pruebas de hipótesis para los contrastes expresadas como H_0: $Q_i = 0$ contra

Tabla 6.5 Análisis de los contrastes definidos en el ejemplo 6.1

	Totales de tratamientos								
	701	*593*	*582*	*641*	*580*	$\hat{Q}_i$	$r\Sigma c_j^2$	$SC(\hat{Q}_i)$	*F*
Q_1	4	−1	−1	−1	−1	408	200	832.32	152.4
Q_2	0	1	1	1	−3	76	120	48.13	8.8
Q_3	0	1	1	−3	1	−168	120	235.20	43.1
Q_4	0	1	−1	0	0	11	20	6.05	1.1 ns

$F_{0.01,1,45} = 7.31$ Suma = 1 121.70

$H_1: Q_i \neq 0$ son altamente significativas ($p < 0.01$), excepto para Q_4. Recuérdese que se rechaza H_0 si $F = SC(\hat{Q})/MCE > F_{0.05,1,45} = 4.06$.

Se concluye que no hay diferencia significativa en los efectos de la glucosa comparados con los de la fructosa; esto responde a la pregunta número cuatro. Para las otras preguntas, la respuesta es que sí hay diferencias significativas desde el punto de vista estadístico.

Intervalos de confianza para contrastes

Otro análisis que permite llegar a los mismos resultados anteriores, es el de intervalos de confianza. Para el primer contraste, basado en medias, su estimativo es $\hat{Q}_1 = 70.1 - (59.3 + 58.2 + 64.1 + 58.0)/4 = 10.2$. Además,

$$\sum_j c_j^2/r = 20/10 = 2,\ MCE = 5.46,\ S(\hat{Q}_1) = \sqrt{5.46 \times 2} = 3.30,\ gl = 45 \text{ y } (t_{0,025,45}) = 2.02.$$

Por tanto, un intervalo del 95% de confianza para Q_1 es $10.2 \pm 2.02 \times 3.30$, es decir, $3.5 \leq Q_1 \leq 16.9$.

En primer lugar, como el intervalo no contiene el cero, puede concluirse que $Q_1 \neq 0$, o sea que hay diferencias en las longitudes ocasionadas por la adición de los azúcares. En segundo lugar, el intervalo indica que la longitud promedio de las secciones de guisantes puede disminuir desde 3.5 hasta un máximo de 16.9 unidades con respecto al control, a causa de la adición de azúcares.

Número diferente de repeticiones por tratamiento

Aunque la mayor eficiencia estadística se obtiene con diseños balanceados, no siempre se logra mantener igual el número de repeticiones por tratamiento. Cuando el tratamiento j-ésimo ha sido repetido r_j veces, las fórmulas para el análisis de contrastes tienen algunas variaciones. Estas nuevas fórmulas se dan a continuación.

La combinación lineal $Q = \sum_j c_j \mu_j$ es un contraste si $\sum_j c_j r_j = 0$.

Los contrastes $Q_1 = \sum_j c_j \mu_j$ y $Q_2 = \sum_j d_j \mu_j$, son ortogonales si $\sum_j r_j c_j d_j = 0$.

El grado de correlación entre Q_1 y Q_2 se mide mediante

$$\rho_{12} = \frac{\sum_j r_j c_j d_j}{\left[\left(\sum_j r_j c_j^2\right)\left(\sum_j r_j d_j^2\right)\right]^{1/2}}.$$

Un estimador de Q basado en medias es $\hat{Q} = \sum_j c_j \bar{y}_{.j}$, con:

$$E\left(\hat{Q}\right) = \sum_i c_j \mu_j,\ V\left(\hat{Q}\right) = \sigma^2 \sum_i c_j^2 / r_j, \text{ y } SC\left(\hat{Q}\right) = \hat{Q}^2 \Big/ \left(\sum_j c_j^2 / r_j\right).$$

Un estimador de Q basado en totales es $\hat{Q} = \sum_j c_j T_j$ con $E\left(\hat{Q}\right) = \sum_j r_j c_j \mu_j$, $V\left(\hat{Q}\right) = \sigma^2 \sum_j r_j c_j^2$ y $SC\left(\hat{Q}\right) = \hat{Q}^2 / \sum_i r_j$.

Un punto importante que se debe tener en cuenta en el análisis de comparaciones múltiples, es que no todas las comparaciones entre medias —planeadas por el experimentador— necesitan ser contrastes. Por ejemplo, la comparación entre los tres azúcares puros no es un contraste, pero puede analizarse calculando directamente su suma de cuadrados, la cual tendrá 2 *gl*, así:

$$SC(\text{azúcares puros}) = \frac{593^2}{10} + \frac{582^2}{10} + \frac{641^2}{10} - \frac{(593 + 582 + 641)^2}{10 + 10 + 10} = 196.87.$$

$$F = \frac{SC(\text{azúcares puros})/2}{MCE} = \frac{98.44}{5.46} = 18.03, p < 0.001.$$

Es necesario aclarar que esta comparación no puede considerarse adicional a los cuatro contrastes anteriores, como una comparación diseñada. Los cuatro contrastes anteriores tienen 1 *gl*; así, en total se tienen 4 *gl*, número correspondiente a los *gl* entre tratamientos del ANOVA —los contrastes se suponen ortogonales—. Si la comparación entre azúcares puros se plantea desde un principio, entonces solo sería posible considerar otros dos contrastes ortogonales para igualar los *gl* del ANOVA y utilizar la prueba F.

Resumiendo, las pruebas F o t pueden usarse cuando se tienen $k - 1$ o menos contrastes ortogonales, o cuando se tiene un conjunto de comparaciones independientes, de modo que tanto las sumas de cuadrados como sus *gl*, no superen la suma de cuadrados y los *gl* entre tratamientos del ANOVA.

6.4 Contrastes no ortogonales y sugeridos por los datos

El investigador podría encontrar que los contrastes con sentido en su experimento fueran no ortogonales. También podría ocurrirle que la información disponible o la teoría existente, no fueran suficientes para formular contrastes antes de realizar el experimento y, entonces, deberá esperar hasta obtener resultados que sugieran el tipo de comparaciones que se deban realizar.

Varios procedimientos permiten analizar contrastes no ortogonales y sugeridos por los datos; entre los más conocidos figuran los de Bonferroni y Scheffé. Para el caso de comparaciones simples, o comparaciones de medias por pares, se tienen los métodos de Tukey, Duncan y otros que se verán posteriormente.

El análisis de estos contrastes incluye pruebas de hipótesis de la forma H_0: $Q_i = 0$ contra H_1: $Q_i \neq 0$, y estimaciones de intervalos de confianza dados por $\hat{Q}_i \pm t_{\alpha/2} S(\hat{Q}_i)$ para $i = 1, 2, \ldots, m$, donde m es el número de contrastes de cualquier tipo que se desea probar.

El análisis estadístico del experimento incluye un conjunto de m pruebas, esto es, una por cada contraste. La validez de cada prueba debe extenderse a dicho conjunto, es decir, el nivel de significación o de error α^* debe referirse al conjunto m y no a cada prueba individual. Este error se interpreta como la probabilidad de encontrar al menos un rechazo erróneo entre la familia de las m hipótesis sobre contrastes, y se conoce como *rata de error experimental*. Se aplica tanto a familias de hipótesis como de intervalos de confianza. En cambio, el nivel de significación usual α para pruebas individuales, es una rata de error por comparación. Hay un problema con estas ratas de error: su relación es extremadamente compleja de calcular y en la mayoría de los problemas prácticos es desconocida, Koopmans (1987).

En general, no se sabe cómo ajustar los niveles de significancia para cada hipótesis de manera que se pueda garantizar un nivel de significancia global preestablecido para la familia completa. Por ejemplo, si se rechaza H_0 con $\alpha = 0.05$ en cada uno de tres contrastes no ortogonales, la probabilidad de que cada prueba conduzca a la aceptación de H_0, cuando en realidad H_0 es correcta, será igual a $(1 - 0.05)^3 = 0.857$, suponiendo independencia entre las pruebas. Así, el nivel de error, dado como la probabilidad de que al menos una de las tres pruebas resulte erróneamente significativa, es $1 - 0.857 = 0.143$ y no 0.05 como se supuso inicialmente para cada prueba.

En conclusión, el nivel de significación para la familia de pruebas es mayor que el nivel individual para cada una, porque las estadísticas muestrales F o t de las diferentes pruebas son dependientes. En realidad, todas ellas están basadas en los mismos datos muestrales y usan la misma varianza (*MCE*) del ANOVA.

6.5 Métodos de Bonferroni y Scheffé

Una aproximación a la solución del problema de multiplicidad de hipótesis, la proporciona una relación de tipo probabilístico conocida como desigualdad de Bonferroni. Esta desigualdad puede establecerse como sigue: sean A_i el evento *rechazar incorrectamente la i-ésima hipótesis nula*, $P(A_i) = \alpha_i$ $(i = 1, 2, \ldots, m)$ y $\bar{A}_i$, el evento complementario de A_i. Entonces:

$$P\left(\bigcap_{i=1}^{m} \bar{A}_i\right) \geq 1 - \sum_{i=1}^{m} P(A_i).$$

Presentada en esta forma, la desigualdad resulta un poco difícil de interpretar, pero puede transformarse, sin que cambie su significado, de la siguiente manera:

$$P\left(\bigcup_{i=1}^{m} A_i\right) \leq \sum_{i=1}^{m} \alpha_i.$$

En palabras sencillas, la desigualdad establece que si varios eventos son de interés, entonces no importa qué tan complejas sean sus interrelaciones: la probabilidad de que ocurra al menos uno de ellos no puede exceder la suma de las probabilidades de los eventos individuales. Aplicada la desigualdad al problema planteado, se concluye que la rata de error experimental no puede ser mayor que la suma de las ratas de error por comparación, esto es que $\alpha^* \leq \Sigma\alpha_i$.

Volviendo al ejemplo de los tres contrastes, si cada uno se prueba con $\alpha = 0.05$, la rata de error experimental α^* para la familia de las tres hipótesis será menor que 0.15. Entonces, $0.05 \leq \alpha^* \leq 0.15$, pero no hay forma de calcular el valor exacto de α^*. Se concluye que la elección de un nivel de significación de 0.05 para hipótesis individuales, no garantiza un nivel de significación del 0.05 para la familia de todas las hipótesis. Al no conocerse o no poderse controlar el error para el experimento, el procedimiento analítico sería inaceptable estadísticamente por falta de validez.

Si se desea una rata de error experimental α^* y se tienen m contrastes o pruebas, el método de Bonferroni consiste en tomar como nivel de significación, para cada una de las pruebas, la probabilidad α^*/m. Por ejemplo, si $\alpha^* = 0.05$ y se han planeado tres contrastes, entonces, cada contraste se probará con $\alpha = 0.05/3 \approx 0.016$. Como el error para la familia completa no excede el valor 3α, puede garantizarse que $\alpha^* \leq 0.05$.

Ejemplo 6.2 Se diseñó un experimento para comparar la efectividad de cinco métodos de almacenamiento en la preservación del contenido de agua de un producto alimenticio. Se hicieron cinco repeticiones por tratamiento en un diseño completamente aleatorizado. Por fallas en el proceso, varias repeticiones se perdieron. Los métodos de almacenamiento y los resultados obtenidos se dan (en porcentajes) en la tabla 6.6.

Se desea comprar materiales —icopor y shopak— y tiempos mediante los contrastes siguientes:

$$Q_1 = (1/3)(\mu_1 + \mu_2 + \mu_3) - (1/2)(\mu_4 + \mu_5).$$

$$Q_2 = (1/2)(\mu_2 + \mu_3) - (1/2)(\mu_4 + \mu_5).$$

Al observar que el primer tratamiento tiene un porcentaje promedio alto, se sugiere además: $Q_3 = \mu_1 - (1/2)(\mu_2 + \mu_3)$.

Cuando se está interesado en un conjunto de contrastes correlacionados pero relativamente pequeño, el método de Bonferroni es con frecuencia más potente que otros procedimientos adoptados para estos casos. Además, este método no requiere un resultado significativo en la prueba F del ANOVA.

Aplicando intervalos de confianza para responder las preguntas sobre comparaciones de métodos de almacenamiento, se tiene una confiabilidad al menos del $(1 - \alpha)100\%$ de que los límites de confianza para los tres contrastes sean todos correctos.

El intervalo de Bonferroni es de la forma $\hat{Q} \pm t_B S\left(\hat{Q}\right)$ con $t_B = t_{\alpha/2,m,\nu}$. Los valores para t_B se obtienen a partir de la tabla de la distribución t mediante interpolaciones lineales, o de tablas construidas para algunos valores de m (véase tabla A.10). Tablas más completas fueron publicadas por Dayton y Schafer (1973) y en Wickens (1989).

Las estimaciones para el contraste Q_1 basado en medias, son:

$$\hat{Q}_1 = (8.08 + 6.63 + 7.25)/3 - (9.13 + 7.10)/2 = -0.795.$$

Tabla 6.6 Pérdida de agua en porcentaje según el tratamiento usado

	Icopor 3 días	*Icopor 10 días*	*Icopor 15 días*	*Shopak 10 días*	*Shopak 15 días*
	7.8	5.4	8.1	7.9	7.1
	8.3	7.4	6.4	9.5	
	7.6	7.1		10.0	
	8.4				
	8.3				
$\bar{y}_{.j}$	8.08	6.63	7.25	9.13	7.10

Tabla 6.7 Tabla del análisis de varianza para los datos del ejemplo 6.2

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Entre métodos	10.90	4	2.72	3.67	0.049
Error	6.69	9	0.74		
Total	17.59	13			

$$\sum_j c_j^2 / r_j = \frac{(1/3)^2}{5} + \frac{(1/3)^2}{3} + \frac{(1/3)^2}{2} + \frac{(1/2)^2}{3} + \frac{(1/2)^2}{1} = 0.448.$$

$$S(\hat{Q}_1) = \sqrt{0.74 \times 0.448} = 0.576. \text{ (0.74 = } MCE \text{ del ANOVA, tabla 6.7).}$$

De la tabla A.10, $t_B = t_{0.025,3,10} = 2.87$; para 9 *gl*, $t_B = 2.90$.

Luego, un intervalo del 95% para Q_1 es $-0.795 \pm 2.9 \times 0.576$, o sea $-2.5 \leq Q_1 \leq 0.9$. Cálculos similares para Q_2 y Q_3 dan los resultados siguientes:

$\hat{Q}_1 = -0.795$ y $-2.5 \leq Q_1 \leq 0.9$.

$\hat{Q}_2 = -1.175$ y $-3.0 \leq Q_2 \leq 0.7$.

$\hat{Q}_3 = 1.140$ y $-1.6 \leq Q_3 \leq 3.9$.

Estos intervalos tienen un coeficiente de confianza común del 95%, o sea que al menos en el 95% de las veces que se repita el experimento, los tres intervalos así calculados serán correctos. Como todos ellos contienen el cero, se concluye que no hay diferencias en cuanto a la preservación del contenido de agua, ni entre los empaques ni entre los tiempos de almacenamiento.

Una observación importante, en la que es necesario llamar la atención en este ejemplo, es el número relativamente pequeño y desbalanceado de repeticiones disponibles. Aunque el diseño estadístico fuera correcto, el no tener la suficiente precaución en el manejo de las unidades experimentales puede llevar a la disminución de eficiencia en el método estadístico. Si el ejemplo fuese una investigación real, deberá pensarse en repetir el experimento.

Si el analista hubiese preferido la prueba t_B de Bonferroni en vez de los intervalos de confianza, habría utilizado la siguiente prueba: rechazar H_0 si $|T_B| \geq t_B$, con $T_B = \hat{Q} / S(\hat{Q})$. Para el contraste Q_3, por ejemplo, se tendría que $T_B = 1.14/0.946 = 1.20 < 2.90$ y no se rechaza H_0. Conclusiones similares se tendrían para los otros dos contrastes; o sea que ninguna de las tres pruebas sería significativa al 5%.

Si el número m de comparaciones para analizar es grande comparado con el número de tratamientos, y si se busca mayor sensibilidad en las pruebas, podría ser preferible el

método de Scheffé al de Bonferroni. Scheffé (1999) ha desarrollado un procedimiento que garantiza una rata de error experimental constante para cualquier familia de contrastes finita o infinita, y para diseños balanceados o desbalanceados. El método fue ideado, específicamente, para comparar contrastes a posteriori sugeridos por los datos y además no exige significancia de la prueba F del ANOVA.

Debido a su gran generalidad, es un método poco potente, pero en muchas situaciones es el recurso más adecuado o incluso el único disponible.

Scheffé demostró que los límites de confianza del tipo $\hat{Q}_1 \pm S \times S(Q)$ son simultáneamente correctos con probabilidad igual a 1 – α. El valor de Scheffé se calcula como: $S = \sqrt{(k-1)F_{\alpha,k-1,n-k}}$. Para los contrastes del ejemplo anterior, este valor común es $S = \sqrt{4F_{0.05,4,9}} = \sqrt{4(3.63)} = 3.81$. Por tanto, los intervalos de confianza serán

$\hat{Q}_1 \pm 3.81\, S(\hat{Q}) = -0.795 \pm (3.81)(0.576)$ o sea, para los otros contrastes los intervalos son:

$$-3.0 \leq Q_1 \leq 1.4.$$

$$-3.6 \leq Q_2 \leq 1.2.$$

$$-2.5 \leq Q_3 \leq 4.7.$$

El aumento en la amplitud de los intervalos por el uso del método de Scheffé (el 30% aproximadamente), es una imprecisión adicional que debe aceptarse para poder garantizar el 95% de confianza en las conclusiones obtenidas sobre los intervalos de confianza, tanto para los tres contrastes como para todas las demás comparaciones que surjan del experimento.

Es obvio que en una investigación nunca se involucran todas las comparaciones concebibles; así que el coeficiente de confianza para la familia finita de contrastes realmente considerados, será mayor que 1 – α. Por esta razón, se ha sugerido que el coeficiente de confianza empleado con Scheffé podría ser menor que el usual, por ejemplo el 90%.

Una alternativa válida cuando todavía no se ha seleccionado un método particular, consiste en aplicar ambos métodos y elegir el que ofrezca mayor precisión. En efecto, esta elección no depende de los datos muestrales.

En algunos textos el problema de comparaciones entre medias se resuelve no por intervalos de confianza, sino directamente comparando el estimado de Q con la amplitud dada por $tS(\hat{Q})$. Así, para probar significancia de Q_1 el procedimiento consiste en rechazar H_0 si $|\hat{Q}_1| > t_B S(\hat{Q}_1)$ según Bonferroni, o $|\hat{Q}_1| > S \times S(\hat{Q}_1)$ según Scheffé. Haciendo los reemplazos correspondientes se tiene:

$|\hat{Q}_1| = 0.795 < 1.67$, luego no se rechaza H_0 (Bonferroni).

$|\hat{Q}_1| = 0.795 < 2.20$, luego no se rechaza H_0 (Scheffé).

6.6 Comparaciones de medias por pares

Esperar a que los datos obtenidos en un experimento influyan en la escogencia de hipótesis que van a probarse con esos mismos datos, no es una buena práctica experimental. Sin embargo, esta falla en el diseño es excusable cuando los tratamientos son clases cualitativas que no están relacionadas o que no es posible diferenciarlas por conocimiento previo. Por ejemplo, varios compuestos experimentales que se analizan como posibles productos farmacéuticos: se supone que la diferenciación

de las naturalezas de los compuestos solo pueden hacerse en forma cualitativa, y que las teorías bioquímicas o fisiológicas no sugieren ningún tipo de contraste. Otro ejemplo podría ser la comparación de artículos, como guantes de trabajo, repuestos automotores, etc., producidos por diferentes empresas.

En casos como estos, el investigador generalmente plantea la comparación de algunos o de todos los pares posibles de medias de tratamientos. Aunque esta familia de contrastes —todos los pares posibles— aparezca como planeada, su análisis corresponde a los métodos para probar efectos sugeridos por los datos. En efecto, si se tienen k tratamientos, el conjunto de comparaciones tendrá $k(k-1)/2$ *pares, un número mayor de* $k-1$, por lo que la familia de contrastes será no ortogonal.

El procedimiento clásico para el análisis de comparaciones por pares fue el de Fisher, conocido como método de la mínima diferencia significativa, *MDS*, y consta de un conjunto de pruebas t entre pares de medias, efectuadas como si las pruebas individuales fueran mutuamente ortogonales. Durante años se han presentado objeciones al uso indiscriminado de este método en comparaciones por pares. La razón fundamental es que el procedimiento no protege adecuadamente al investigador de obtener conclusiones erróneas.

Supóngase cinco tratamientos en los cuales todas las medias μ_j sean iguales. Existen diez comparaciones posibles entre pares de medias, y la probabilidad de que por lo menos una de las diez diferencias exceda la *MDS* será siempre mayor que 0.05. En efecto, puede probarse que la probabilidad está cercana a 0.29. Con diez tratamientos, la probabilidad de encontrar al menos una diferencia significativa es aproximadamente igual a 0.63 (Snedecor y Cochran, 1984).

Tratando de resolver este problema, se han planteado muchos procedimientos estadísticos que buscan darle validez a las comparaciones por pares, en el sentido de garantizar una rata de error experimental constante. Las páginas siguientes estarán dedicadas al estudio de algunos de los métodos de comparaciones por pares más comúnmente usados, pues hasta la fecha no existe acuerdo, al menos entre los estadísticos, sobre qué método debería usarse ante una situación experimental más o menos rutinaria.

6.7 Método de Fisher, MDS

En general, las pruebas de comparaciones por pares tienen los mismos supuestos requeridos en el ANOVA: normalidad de las poblaciones y homogeneidad de varianzas. Además, aunque no sea un requisito teórico, estas pruebas se realizan sólo si el resultado del análisis de varianza es significativo (Berhardson, 1975).

El método de Fisher se realiza en dos pasos: el primero es el ANOVA estándar generalmente al 5%. Si se acepta H_0, el proceso termina y se concluye que no hay suficiente evidencia en los datos para sustentar diferencias entre medias poblacionales. Si se rechaza H_0, el segundo paso consiste en aplicar la prueba t de diferencia de medias a cada par de medias con el mismo α del ANOVA. El problema de la rata de error experimental no es tan serio en este caso. Carmer y Swanson (1973) han encontrado que su valor puede estar cercano al 5%, de ahí que se le conozca también como la prueba de Fisher protegida.

Si la prueba t es no significativa, las dos medias se asignarán a un mismo grupo; si es significativa, estas se ubicarán en grupos diferentes. Mediante este proceso se logra una agrupación de medias que proporciona información rápida y completa. La mínima diferencia significativa que sirve de comparación se calcula como:

$$MDS = t_{\alpha/2,\nu}\sqrt{MCE\left(1/r_i + 1/r_j\right)}.$$

donde r_i y r_j son las repeticiones de los dos tratamientos comparados.

Ejemplo 6.3 Un investigador en farmacia quiso comparar diferencias en los tiempos de desintegración de cuatro tipos de pastillas, A, B, C y D. Para ello seleccionó seis pastillas al azar de cada tipo y midió los tiempos de desintegración; así obtuvo los datos de la tabla 6.8 (adaptado de Anderson y MacLean, 1974). La tabla 6.9 muestra el análisis de varianza correspondiente.

La prueba F permite concluir que hay diferencias significativas entre los tiempos promedios de desintegración de las pastillas.

Una forma de reconocer estas diferencias consiste en listar todos los pares de medias posibles, comenzando por aquellos con las medias más extremas; calcular la diferencia de medias de cada par y luego comparar estas diferencias con el valor crítico, obtenido como

$$MDS = 2.086\sqrt{3.4(1/6 + 1/6)} = 2.22.$$

Las medias ordenadas son:

A	*C*	*D*	*B*
5.0	5.7	6.8	8.7

Diferencias entre medias mayores que 2.22 son significativas al 5%, así:

$B - A = 3.7 > 2.22$, diferencia significativa

$B - C = 3.0 > 2.22$, diferencia significativa.

$D - A = 1.8 < 2.22$, diferencia no significativa.

$B - D = 1.9 < 2.22$, diferencia no significativa.

Una forma gráfica para presentar los resultados fue sugerida por Duncan; esta consiste en colocar a escala las medias ordenadas de menor a mayor y subrayar con la misma línea aquellas medias que forman grupos homogéneos. Por

Tabla 6.8 Tiempos de desintegración (en segundos) para cuatro pastillas diferentes

	A	*B*	*C*	*D*
	6	10	3	10
	2	8	7	4
	5	11	6	6
	4	7	4	6
	6	7	8	7
	7	9	6	8
Totales	30	52	34	41
Medias	5.0	8.7	5.7	6.8

Tabla 6.9 Tabla del ANOVA para los datos del ejemplo 6.3

Fuente de variación	SC	gl	MC	F	p
Entre tipos de pastillas	46.46	3	15.5	4.06	0.013
Error	67.50	20	3.4		
Total	113.96	23			

supuesto, dos medias no subrayadas con la misma línea son diferentes. Para el ejemplo anterior véase la figura 6.1.

De este gráfico puede concluirse lo siguiente:

1. La pastilla B tiene un tiempo promedio de desintegración significativamente mayor que el de las pastillas A y C.
2. No hay diferencia entre las pastillas A y C respecto a su tiempo promedio de desintegración.
3. No hay suficiente evidencia en los datos obtenidos para determinar a cuál de los dos grupos pertenece la media de la pastilla D, es decir, se ha cometido al menos un error de tipo II.

Algunos paquetes estadísticos presentan comparaciones por pares, ya sea mediante una tabla $k \times k$ o separando los grupos homogéneos por columnas e identificándolos con asteriscos. Para el ejemplo anterior, la presentación sería:

		Grupos homogéneos	
Pastillas	*Medias*	*1*	*2*
A	5.0	*	
C	5.7	*	
D	6.8	*	*
B	8.7		*

Grupos traslapados, como en este ejemplo, resultan claramente indeseables dada la dificultad para su interpretación, pero aparecen en el análisis y muchas veces son inevitables. El traslape podría desaparecer si se usa un procedimiento de mayor sensibilidad o si se refina el método, por ejemplo, usando medias recortadas cuando se han reconocido algunos valores extremos. Cabe recordar que para ganar potencia en una prueba deberá repetirse el estudio con un número mayor de repeticiones.

Cuando el número de repeticiones es desigual, teóricamente sería necesario calcular una *MDS* para cada par de medias; en la práctica, se calculan las *MDS* máxima y mínima. Diferencias mayores que la máxima *MDS* serán significativas y diferencias menores que la mínima serán no significativas. De esta manera, solo se calcula la *MDS* para aquellas diferencias que estén comprendidas entre las *MDS* máxima y mínima.

Ejemplo 6.4 En el ejemplo 4.2, sobre tiempos de coagulación de la sangre, se encontró que había diferencias altamente significativas entre las dietas. Para hallar estas diferencias se aplicará el método de comparaciones de Fisher.

Para calcular las *MDS* máxima y mínima se obtiene primero el valor:

$t_{a/2,v}\sqrt{MCE} = 2.086\sqrt{5.6} = 4.94$. Luego,

MDS máx. $= 4.94\sqrt{1/4 + 1/6} = 3.2$ y

MDS mín. $= 4.94\sqrt{1/8 + 1/6} = 2.7$.

A	C	D	B
5.0	5.7	6.8	8.7

Figura 6.1 Gráfico de Duncan para presentar los resultados del ejemplo 6.3

D_1 D_4	D_2	D_3
61	66	68

Figura 6.2 Gráfico de Duncan para el ejemplo 6.4

Las diferencias $D_3 - D_1$, $D_3 - D_4$, $D_2 - D_1$ y $D_2 - D_4$ son mayores que 3,2 y, por tanto, significativas; mientras que las diferencias $D_3 - D_2$ y $D_4 - D_1$ son menores que 2.7, o sea, no significativas. En este ejemplo no hubo diferencias intermedias.

La figura 6.2 permite diferenciar claramente dos grupos homogéneos: uno compuesto por las dietas D_1 y D_4 y el otro por las dietas D_2 y D_3. El tiempo de coagulación del primer grupo es significativamente menor que el del segundo grupo.

El método de la *MDS* ha sido usado correcta e incorrectamente en la literatura. Según Steel y Torrie (1997), los usos incorrectos más comunes se presentan cuando se aplica en comparaciones sugeridas por los datos y cuando se comparan todos los pares posibles. El método es incorrecto en el sentido de que la rata de error experimental es desconocida, aunque el método con protección —*F* significativa— da un α^* alrededor de 0.05. Cuando se especifican previamente los pares que van a compararse y estos son ortogonales, es válido el uso de la *MDS*.

6.8 Método de Tukey, DHS

El método de Tukey utiliza la distribución del rango *estudentizado* para fijar el valor crítico con el cual se comparan las diferencias entre las medias. Sea el conjunto de k medias $\bar{y}_{.1}, \bar{y}_{.2}, \cdots \bar{y}_{.k}$ con distribución Normal, y $S(\bar{y}_{.j})$ un estimador de la desviación estándar de $\bar{y}_{.j}$.

El cociente $q(k,\nu) = (\text{máx}\,\bar{y}_{.j} - \text{mín}\,\bar{y}_{.j}) / S(\bar{y}_{.j})$ se llama rango estudentizado y su distribución se encuentra tabulada (véase tabla A. 12) para valores de α, k y ν, donde ν son los *gl* de la *MCE*.

El valor crítico de Tukey se conoce como diferencia significativa honesta y se calcula como $DHS = q_\alpha(k,\nu)\sqrt{MCE/r}$. Entonces, se declara significativa toda diferencia de medias mayor que *DHS*; el método garantiza un α común para todas las comparaciones.

El método de Tukey aplicado al ejemplo 6.3 produce los mismos resultados obtenidos con el método de Fisher. Para $\alpha = 0.05$, $k = 4$, y $n - k = 20$, se lee $q_{0.05}(4,20) = 3.96$ y $DHS = 3.96\sqrt{3.4/6} = 2.98$.

Se observa que al cambiar el valor crítico de 2.22 a 2.98, las conclusiones no varían. No siempre ocurre así, ya que el método de Tukey origina una diferencia mayor que la *MDS* para establecer significancia y, por tanto, serán detectadas menos diferencias reales. Esta pérdida de precisión es el riesgo que se corre si se quiere obtener una tasa de error experimental no mayor que α.

El método de Tukey se aplica generalmente a comparaciones por pares dando lugar a un valor crítico menor que el de Scheffé o el de Bonferroni, que incluyen dichas comparaciones como un caso particular.

Puede ocurrir que el investigador esté interesado en el tipo de pastilla con el menor tiempo de desintegración; entonces, selecciona uno de los tipos A o C. Un intervalo del 95% de confianza para el tiempo de desintegración es $\bar{y}_{.j} \pm t_{\alpha/2,\nu}\sqrt{MCE/r_j}$. Para la pastilla A se tiene $5.0 \pm 2.086\sqrt{3.4/6}$, o sea, $3.4 \le \mu_A \le 6.6$.

Una alternativa má s práctica consiste en promediar las dos medias homogé-

neas para obtener $\bar{y} = (\bar{y}_A + \bar{y}_C)/2$ = (5.0 + 5.7)/2 = 5.35 y calcular el intervalo con $r_j = 12$; así $5.35 \pm 2.086\sqrt{3.4/12} = (4.2, 6.5)$.

Las propiedades de la *DHS* son válidas únicamente para diseños balanceados, pero con algunas modificaciones, el método también puede usarse en el caso de diferentes r_j siempre que el desbalance no sea grande. La *DHS* modificada es $q(k,v)\sqrt{MCE/r'}$, con $r' = \text{mín}\ (r_i, r_j)$. El grado de desbalance se mide mediante el cociente $u = \text{máx}\ r_j / \text{mín}\ r_j$. Entonces, según Spjøtvoll y Stoline (1973), se prefiere Tukey a Scheffé para $k \geq 15$ con cualquier u; para $6 \leq k < 15$ cuando $u \leq 2$ y para $k < 6$ cuando $u \leq 1.2$.

En el ejemplo 6.4 no es recomendable usar Tukey modificado puesto que $k = 4$ y $u = 2$. Sin embargo, Kramer (1956) recomienda a Tukey con $DHS = q(k,v)\sqrt{(MCE/2)(1/r_i + 1/r_j)}$; otros autores han aplicado esta sugerencia con resultados excelentes. Utilizando esta última *DHS* para el ejemplo de las dietas, los resultados son iguales a los obtenidos con el método de Fisher, como se ve a continuación:

$D_3 - D_1 = 7 > 4.28$, diferencia significativa,

$D_3 - D_4 = 7 > 3.58$, diferencia significativa,

$D_3 - D_2 = 2 < 3.82$, diferencia no significativa,

$D_2 - D_1 = 5 > 4.28$, diferencia significativa,

$D_2 - D_4 = 5 > 3.58$, diferencia significativa,

$D_4 - D_1 = 0 < 4.06$, diferencia no significativa.

6.9 Método de Student-Newman-Keuls, SNK

El SNK es un método que permite investigar de manera secuencial todos los pares posibles de medias, garantiza una rata de error α constante y tiene buena potencia. Si se le exige que la prueba *F* del ANOVA sea significativa, el método puede tener excelente control sobre la rata de error de tipo I y resulta especialmente apropiado cuando en la investigación hay involucrados varios factores con sus interacciones.

Para realizar la prueba se ordenan las medias y se prepara una tabla de diferencias de medias de modo que se forme un arreglo triangular, como se muestra en el ejemplo 6.5. Luego se calculan rangos críticos estudentizados según el número de medias incluidas en el grupo de comparación, de aquí que se le llame prueba de rango múltiple.

Los rangos mínimos significativos se calculan mediante los productos:

$$W_p = q_\alpha\,(p, gl)\sqrt{(MCE/r)},\ \text{con}\ p = k, k-1, \ldots, 2.$$

Sean $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{k-1}, \bar{x}_k$ las medias ordenadas de menor a mayor. El método SNK inicia, como el de Tukey, comparando las medias más distantes. Si la diferencia $\bar{x}_k - \bar{x}_1$ es menor que W_k, —diferencia no significativa—, se declara homogéneo el conjunto de medias y no se hacen más comparaciones; si la diferencia es significativa, se con-

cluye que $\bar{x}_1$ y $\bar{x}_k$ son diferentes. Luego se prueban las diferencias $\bar{x}_k - \bar{x}_2$ y $\bar{x}_{k-1} - \bar{x}_1$ para lo cual se usa el rango W_{k-1}. Si alguna de estas diferencias es no significativa, entonces se supone homogeneidad en ese grupo y no se hacen más pruebas dentro de este. El procedimiento continúa hasta que todas las diferencias sean no significativas, o hasta terminar con todas las comparaciones.

Ejemplo 6.5 En una granja experimental, se estudió la producción de lana en cada esquileo de ovejas durante seis periodos a intervalos de seis meses (adaptado de Das y Giri, 1986). Los promedios en onzas por esquileo fueron, en orden creciente de rendimiento, los siguientes:

A	*B*	*C*	*D*	*E*	*F*
16.91	20.70	20.81	24.50	26.12	31.62

El diseño fue de tipo factorial, y en el análisis estadístico se encontró que una estimación para la varianza era $MCE = 25.78$, con $10\,gl$ y $r = 24$. Además, la prueba F del ANOVA fue significativa con $p < 0.01$. Con esta información se construye la tabla 6.10

Las diferencias entre las medias se dan en una tabla triangular (véase tabla 6.11) y los grupos de comparación aparecen en la columna de la izquierda. Las diferencias situadas sobre la misma diagonal —de izquierda a derecha— se prueban con el mismo valor de p.

Las comparaciones se efectúan siguiendo las diagonales de la tabla de diferencias, así:

Para $p = 6$; $F - A = 14.71 > 5.09 = W_6$; luego A y F son significativamente diferentes.

Para $p = 5$; $F - B = 10.92 > 4.82$, diferencia significativa.

$E - A = 8.21 > 4.82$, diferencia no significativa

Para $p = 4$; $F - C = 10.81 > 4.48$, diferencia significativa.

$E - B = 4.42 < 4.48$, diferencia no significativa

$D - A = 7.59 > 4.48$, diferencia significativa.

Las medias correspondientes a B, C, D y E, forman un grupo homogéneo y, por tanto, no se comparan otras medias dentro de este grupo.

Para $p = 3$: $F - D = 7.12 > 4.02$, diferencia significativa

$C - A = 3.90 < 4.02$, diferencia no significativa.

Las medias en el grupo A, B y C son homogéneas.

Para $p = 2$: $F - E = 6.50 > 3.26$, diferencia significativa.

El gráfico de Duncan para estas medias se observa en la figura 6.3.

Se concluye: 1) el valor promedio del esquileo F fue mayor que el de los demás, 2) el promedio del esquileo A fue menor que el de los esquileos D, E y F y 3) los esquileos B y C

Tabla 6.10 información obtenida para el ejemplo 6.5

p	$q_{0.05}(p,10)$	W_p
2	3.15	3.26 → *MDS*
3	3.88	4.02
4	4.33	4.48
5	4.65	4.82
6	4.91	5.09 → *DHS*

Tabla 6.11 Tabla de diferencias para el ejemplo 6.5

p		*A*	*B*	*C*	*D*	*E*
6	*F*	14.71	10.92	10.81	7.12	6.50
5	*E*	8.21	4.42	4.31	0.62	
4	*D*	7.59	3.80	3.69		
3	*C*	3.90	0.11			
2	*B*	3.79				

A 16.91 — BC 20.70 20.81 — D 24.50 — E 25.12 — F 31.62

Figura 6.3 Gráfico de Duncan para el ejemplo 6.5

fueron similares entre sí, pero los datos no ofrecen suficiente evidencia para diferenciarlos de A y tampoco de D y E, que también fueron similares entre sí.

Para una interpretación más real del ejemplo, los esquileos en orden cronológico fueron: uno para A, dos para B, tres para D, cuatro para F, cinco para E y seis para C.

Puede ocurrir en algunas investigaciones, que después de declarar una diferencia entre medias no significativa, se encuentre una diferencia significativa entre otras medias dentro del mismo grupo homogéneo. Esta última diferencia debe declararse no significativa con el fin de evitar contradicciones lógicas. Este es precisamente el caso con la diferencia $E - C = 4.31 > 4.02$, cuando la diferencia $E - B = 4.42$ había sido declarada no significativa. Por esto no se deben continuar las comparaciones en aquellos grupos declarados homogéneos.

Para grupos desiguales, el método SNK no es exacto, pero produce buenos resultados si se aplica con las reformas recomendadas por Kramer para el método de Tukey.

Aplicando el método de Tukey con $DHS = 5.09$ al ejemplo 6.5, se obtienen exactamente las mismas conclusiones anteriores. El método de Fisher, en cambio, resulta más sensible ya que la *MDS* (MDS = 3.26) es mucho menor que 5.09. El gráfico correspondiente se observa en la figura 6.4.

6.10 Método de Duncan, RMD

Otro método de comparaciones por pares es el del rango múltiple de Duncan, RMD. Es un método que proporciona mejor potencia que los anteriores, pero no protege al investigador con respecto a la probabilidad de error de tipo I.

El procedimiento es similar al SNK en cuanto a que también usa rangos múltiples a partir de una amplitud estudentizada, pero se diferencia porque utiliza un nivel de significación variable $\alpha_p = 1 - (1 - \alpha)^{p-1}$, donde α es el nivel de significación inicial y p el número de medias en el grupo de comparación.

Los rangos estudentizados de Duncan $r_\alpha\ (p, gl)$ deben leerse en las tablas proporcionadas por Duncan (véase tabla A.12), y los rangos mínimos significativos se calculan con $R_p = r_\alpha\,(\mathrm{p}, gl)\sqrt{MCE/r}$.

Para el ejemplo de la producción de lana, se construye la tabla 6.12.

Al efectuar las comparaciones respectivas se llega a las mismas conclusiones obtenidas con el método de Fisher.

El método RMD se conoce en la literatura como la nueva prueba de amplitud

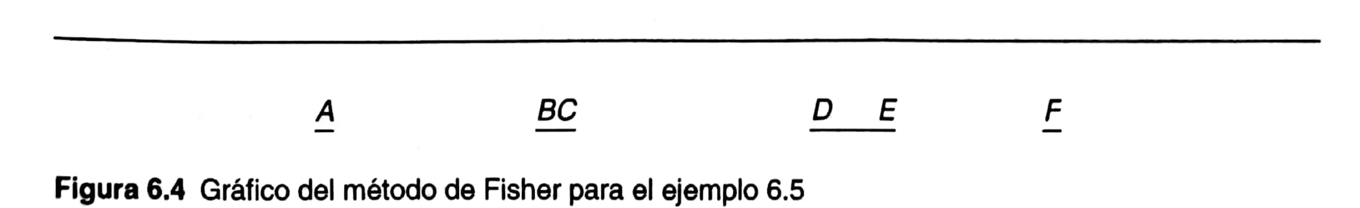


Figura 6.4 Gráfico del método de Fisher para el ejemplo 6.5

Tabla 6.12 Datos del método RMD para el ejemplo 6.5

p	$r_{0.05}$ (p, 10)	R_p
2	3.15	3.26
3	3.30	3.42
4	3.37	3.49
5	3.43	3.55
6	3.46	3.58

múltiple de Duncan, Duncan (1955), porque este mismo autor había propuesto otra prueba en 1951.

6.11 Método de Dunnett: comparaciones con un control

En algunos experimentos, el objetivo fundamental es la comparación de varios tratamientos con un tratamiento estándar, llamado tratamiento control. El enfoque de este problema no puede ser el mismo de comparaciones múltiples por pares, ya que solo se está interesado en $k - 1$ comparaciones. Además, las comparaciones con el control están mutuamente correlacionadas, por lo que no pueden considerarse como comparaciones planeadas.

Dunnett (1955) desarrolló un procedimiento para resolver este problema con la ventaja de que mantiene constante la rata de error experimental; es similar a las pruebas anteriores de comparaciones por pares, pero los valores críticos se leen en tablas construidas por el mismo autor (véase tabla A.13). El valor crítico de Dunnett se calcula como $D = t_D\sqrt{2\,MCE/r}$, con $t_D = t_\alpha(k-l, gl)$. El valor de t_D se lee en las tablas dependiendo de si la hipótesis es unilateral o bilateral.

Ejemplo 6.6 En un programa experimental sobre la resistencia a la rotura de diafragmas, se compararon seis tipos de caucho con un tipo estándar. Los datos obtenidos se muestran en la tabla 6.13.

Con cada uno de los seis tipos más el control, se hicieron $n = 4$ repeticiones y se obtuvo una $MCE = 9.0$ con 21 gl. En las tablas de Dunnett se lee $t_{0.05}$ (6, 20) = 2.80 (valor aproximado ya que la tabla no trae el valor para 21 gl). Así, el valor crítico es $D = 2.80\sqrt{2 \times 9.0/4} = 5.94$. Cualquier diferencia con el control mayor que 5.94 en valor absoluto puede considerarse estadísticamente significativa (véase tabla 6.14).

Solamente el tipo de caucho E tiene una diferencia significativa con el control.

Cuando se comparan varios tratamientos con un control, el diseño óptimo se obtiene al hacer un número mayor de repeticiones del control que de cada uno de los otros tratamientos; la relación deberá ser $r_c / r_j \approx \sqrt{k-1}$.

En el ejemplo anterior, para el control deberían hacerse al menos diez repeticiones.

Tabla 6.13 Datos del ejemplo 6.6

Tipo de caucho:	*Estándar*	*A*	*B*	*C*	*D*	*E*	*F*
Medias	63	62	67	65	65	70	60
Varianzas	9.2	8.7	8.8	9.8	10.2	8.3	8.0

Tabla 6.14 Datos del ejemplo 6.6 y diferencias con el control

Tipo de caucho	*Control*	*A*	*B*	*C*	*D*	*E*	*F*
Medias	63	62	67	65	65	70	60
Diferencias		+1	–4	–2	–2	–7	+3

El valor crítico será $D = t_D\sqrt{MCE\left(1/r_j + 1/r_c\right)}$. Los valores tabulados para t_D son válidos solo para $r_j = r_c$. Los números entre paréntesis que acompañan los valores tabulados son factores de ajuste para el caso desbalanceado, tal que $1 < r_c/r_j < 7$. Este factor de ajuste se multiplica por $1 - r_j/r_c$ y da el porcentaje de incremento requerido en el valor tabulado. Si en el ejemplo de los diafragmas se hubiesen tomado $r_c = 10$ repeticiones $(10/4 = 2.5 < 7)$, el factor de ajuste sería 3.9 (1 − 0.4) = 2.34 y el valor t_D ajustado sería 2.80(1 + 0.0234) = 2.87. Por tanto, $D = 2.87\sqrt{9.0(1/4 + 1/10)} = 5.09$, suponiendo los mismos valores del ejemplo.

Usando la prueba t, y suponiendo comparaciones independientes, se tendría un valor $D = 2.08\sqrt{9.0(1/4 + 1/10)} = 3.69$ y se concluiría que el caucho tipo B también es diferente del estándar, lo que sería una conclusión errónea desde el punto de vista estadístico.

Las tablas de Dunnett suponen homogeneidad de varianzas, pero los factores de ajuste pueden adaptarse al caso de heteroscedasticidad. Gill (1977, vol. 1) desarrolló un procedimiento similar, más general que el de Dunnett, que puede usarse en experimentos con varianzas marcadamente heterogéneas.

6.12 Método de Dunn: comparaciones por pares usando rangos

En forma similar a las comparaciones de medias por pares, puede efectuarse la comparación por rangos promedio, siempre que en la prueba de Kruskall y Wallis se haya rechazado la hipótesis nula. Las comparaciones pueden llevarse a cabo siguiendo un procedimiento paralelo al de Tukey para grupos iguales, o al de Student-Newman-Keuls para grupos desiguales (Zar, 1999).

Un procedimiento diferente fue propuesto por Dunn (1964) con las ventajas de que admite muestras desiguales y los valores críticos se calculan con base en la distribución Normal. Según la prueba de Dunn, dos tratamientos son diferentes si:

$$\left|\bar{R}_i - \bar{R}_j\right| \geq Z_{\alpha/k(k-1)}\sqrt{\left(1/r_i + 1/r_j\right)n(n+1)/12}.$$

Ejemplo 6.7 En el ejemplo 5.5 se concluyó que la pérdida de peso era diferente para los cinco tipos de carne. Se verá ahora cuáles son esas diferencias.

Para $\alpha = 0.05$ se tiene que $Z_{0.05/20} = Z_{0.0025} = 2.81$. Por otra parte, el error estándar es $\sqrt{(1/5 + 1/5)25 \times 26/12} = 4.655$. Por tanto, el valor crítico es 2.81 × 4.655 = 13.08.

Las comparaciones entre los rangos promedios pueden establecerse a partir de estos valores si se ordenan en la siguiente forma:

Suma de rangos (R_j):	21	44	56	91	113
Medias de rangos ($\bar{R}_j$)	4.2	8.8	11.2	18.2	22.6
	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_5$	$\bar{R}_3$	$\bar{R}_4$

$R_4 - R_1 = 18.4 > 13.1$, diferencia significativa.

$\bar{R}_4 - \bar{R}_2 = 13.8 > 13.1$, diferencia significativa.

$\bar{R}_3 - \bar{R}_1 = 14.0 > 13.1$, diferencia significativa.

$\bar{R}_4 - \bar{R}_5 = 11.4 < 13.1$, diferencia no significativa.

$\bar{R}_3 - \bar{R}_2 = 9.4 > 13.1$, diferencia no significativa.

$\bar{R}_5 - \bar{R}_1 = 7.2 < 13.1$, diferencia no significativa.

Las únicas diferencias claramente definidas se presentan entre el tocino, que pierde menos peso que la milanesa de cerdo y la milanesa de res, y entre la costilla y la milanesa de res.

Esta prueba mantiene constante la rata de error experimental, por lo que resulta muy conservadora; es decir, mantiene alto grado de protección cuando H_0 es verdadera. Al contrario, si H_0 es falsa, este procedimiento detecta con menos frecuencia la significancia entre las diferencias de los tratamientos. Es lo que ocurre en el ejemplo anterior, en el cual se tienen tres grupos completamente traslapados (véase figura 6.5).

El método de Dunn es aplicable cuando el número de repeticiones por grupo es grande (10 o más). Lee y Wang (2003) trae otros procedimientos y sus respectivas tablas, para comparar rangos cuando los grupos son pequeños, iguales o desiguales.

La dificultad en el reconocimiento de diferencias significativas es mayor a medida que aumenta el número k de tratamientos. Nuevamente, tiene sentido e importancia la recomendación de diseñar experimentos con un número adecuado de repeticiones por tratamiento, para evitar las dificultades anteriores.

Algunas conclusiones sobre el uso y características de los métodos de comparaciones múltiples son:

1. Usar contrastes diseñados y preferiblemente ortogonales, con el fin de maximizar la sensibilidad de las pruebas.

2. Si se requiere comparar todos los pares de medias y el error de tipo I es grave y debe controlarse, úsese Tukey o SNK que protegen contra este error, teniendo en cuenta que el de Tukey es más conservador mientras que el SNK es más liberal con respecto al error α y, además, más potente. Los métodos de Duncan y Fisher son mucho mejores en cuanto a potencia, pues detectan diferencias reales con mayor frecuencia, pero a menudo distorsionan la importancia que debe darse a la evidencia de los datos.

3. Los métodos de Scheffé y Bonferroni no son recomendables para comparaciones por pares por su baja potencia. En general, producen valores críticos más grandes e intervalos de confianza mucho más amplios, con lo cual se dejan de reconocer diferencias verdaderas entre la medias.

4. En general, las pruebas de comparaciones múltiples tienen los mismos supuestos del análisis de varianza. La prueba de Tukey parece ser la más robusta con respecto a desviaciones de las suposiciones, sin embargo, la robustez de estas pruebas no ha sido aún suficientemente estudiada.

5. El problema de desbalanceo en los tratamientos hace que las pruebas pierdan potencia y robustez, pero si este no es muy drástico, los procedimientos anteriores todavía son válidos.

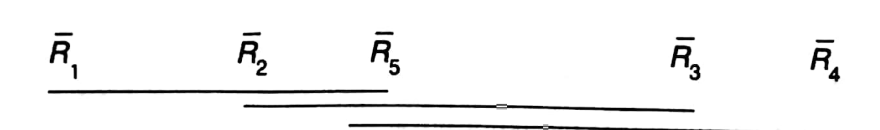


Figura 6.5 Gráfico de Duncan para el ejemplo 6.7

6. Cuando los supuestos llegan a ser insostenibles, es preferible usar un método no paramétrico seguido de una prueba de comparaciones múltiples, igualmente, no paramétrica.

7. Podría ocurrir que la hipótesis H_0 del análisis de varianza fuese rechazada y que la prueba de comparaciones por pares de Tukey, u otra de ellas, no detectara ninguna diferencia significativa entre las medias. Esta ocurrencia parece no ser frecuente, pero si se presenta en una investigación, la explicación es que la prueba F del análisis de varianza es más potente que la prueba de comparaciones múltiples.

Con el nombre común de comparaciones múltiples existe en la literatura un número grande de pruebas, cada una proponiendo una solución a problemas específicos. Fuera de las ya citadas se tienen, por ejemplo, las pruebas de Dunn-Sidak y de Hochberg que son modificaciones a Bonferroni, la prueba Hsu (MCB: comparaciones múltiples con el mejor) que compara los tratamientos con el mejor de ellos (máximo o mínimo) y otras como la de Waller-Duncan, la de Ryan-Einot y la de Peritz.

Los procedimientos de comparaciones múltiples han sido analizados, en cuanto a su comportamiento relativo por varios autores como Dunnett (1970), Miller (1981) y Hsu (1996). Jones (1984), Saville (1990) y Lindman (1992) debaten la aplicación de estas técnicas llegando a recomendar el no uso de algunas de ellas en estudios experimentales.

Ejercicios

6.1 Comparar el promedio de la semana cero con el promedio de las semanas restantes en el ejercicio 4.9.

6.2 Se diseñó un experimento para comparar cinco medios de cultivo con respecto a la capacidad de crecimiento de células fibroblásticas de tejidos de ratones. Para cada medio se tomaron cinco frascos con igual número de células depositadas en cada uno; y después de siete días se determinó la proteína celular total. Los resultados (en microgramos de proteína nitrogenada) se dan en la siguiente tabla.

Frascos	*Medios de cultivo*				
	M_1	M_2	M_3	M_4	M_5
1	102	103	103	108	113
2	101	105	107	101	117
3	100	100	105	104	106
4	105	108	105	106	115
5	101	102	106	104	116

El medio M_1 es un control, los medios M_2, M_3 y M_4 tienen compuestos similares, y el medio M_5 es un control alterno.

a. Comparar los dos controles.

b. Comparar el promedio de los controles con el promedio de los otros medios.

c. Comparar los tres medios M_2, M_3 y M_4 entre sí. Obsérvese que esta comparación no es un contraste.
d. Formular un contraste que sea ortogonal con los dos primeros.
e. Comprobar que el contraste que compara los primeros cuatro medios con el último no es ortogonal con los contrastes en *a* y en *b*.
f. Probar el contraste en *e*.

6.3 Se diseñó un experimento para comparar las cantidades de agua que libera una hoja de árbol laurel cerezo. El factor fue la superficie de la hoja con cuatro tratamientos: hoja sin cubrir, cubriendo la parte superior, cubriendo la parte inferior y cubriendo ambas superficies. Una muestra de 16 hojas se asignó al azar a los cuatro tratamientos en grupos iguales. Cada hoja se pesó y se le midió la superficie. Al cabo de tres días de aireación se pesaron nuevamente. La pérdida de peso (en mg/cm^2) se da en la siguiente tabla.

Hojas	*Superficie cubierta*			
	Ninguna	*Superior*	*Inferior*	*Ambas*
1	86	41	25	13
2	108	44	35	11
3	118	40	37	13
4	79	52	26	13

a. Comparar la pérdida de agua en ambas superficies.
b. Comparar superficies revestidas con la superficie sin revestir.
c. Comparar el revestimiento por una superficie con el revestimiento por ambas superficies.
d. ¿Son los tres contrastes anteriores, ortogonales entre sí?
e. Describir las conclusiones de este experimento y decir si pueden darse con una rata de error experimental común.

6.4 Los registros siguientes contienen los pesos, en libras, al nacer de ocho camadas de cerdos Poland China (Scheffé, 1999).

Camadas							
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
2.0	3.5	3.3	3.2	2.6	3.1	2.6	2.5
2.8	2.8	3.6	3.3	2.6	2.9	2.2	2.4
3.3	3.2	2.6	3.2	2.9	3.1	2.2	3.0
3.2	3.5	3.1	2.9	2.0	2.5	2.5	1.5
4.4	2.3	3.2	3.3	2.0		1.2	
3.6	2.4	3.3	2.0	2.1		1.2	
1.9	2.0	2.9	2.1				
3.3	1.6	3.4					
2.8		3.2					
1.1		3.2					

a. Suponer que las camadas 1, 3 y 4 provienen de un solo reproductor y el resto de otro reproductor. Comparar reproductores.
b. Comparar madres del primer reproductor, esto es, camadas 1, 3 y 4.
c. Comparar las camadas grandes con las pequeñas. Esta comparación se planeó después de conocer los datos.

6.5 Comparar envolturas en el ejercicio 4.2 usando el método de Tukey.

6.6 Comparar los diferentes bosques en el ejercicio 4.3 usando el método de Tukey. Aplicar también el método de Fisher y comparar resultados de los dos métodos.

6.7 Comparar el consumo promedio de larvas por instar en el ejercicio 4.7 usando el método SNK.

6.8 Comparar estaciones de muestreo en el ejercicio 4.4 usando el método de Duncan.

6.9 Comparar años de fabricación de los automóviles en el ejemplo 5.6. Usar el método SNK u otro que usted considere válido.

6.10 Comparar envolturas en el ejercicio 5.6 usando el método de los rangos de Dunn.

6.11 Comparar los grupos asmáticos en el ejercicio 5.8 usando el método de los rangos de Dunn.

6.12 Comparar los pesos al nacer de niños de madres fumadoras con las madres control del ejercicio 4.10. Usar el método de Dunnett.

7

Diseño de bloques aleatorizados

7.1 Diseño de bloques completos aleatorizados

El objetivo central de un diseño estadístico es la reducción del error experimental. Ahora, bien, bajo condiciones experimentales normales, el tamaño de este error depende de la variabilidad de las unidades experimentales; por tanto, si el experimento se lleva a cabo con unidades experimentales homogéneas, puede garantizarse que la media cuadrática del error será relativamente pequeña y, tanto las pruebas como las estimaciones subsiguientes, tendrán mayor sensibilidad y precisión.

En general, en el diseño completamente aleatorizado, DCA, el número de unidades experimentales, *ue*, es grande y resulta difícil o costoso obtener *ue* homogéneas. En estos casos se recomienda cambiar el DCA por otro diseño más eficiente respecto al control del error experimental. El diseño más recomendado y utilizado es el que considera la idea de estratificación, la cual consiste en agrupar las *ue* en subconjuntos homogéneos y luego aplicar aleatoriamente los tratamientos a cada subconjunto.

El grupo o subconjunto de unidades se llama *bloque*, palabra tomada de la experimentación agrícola donde fue usada para reconocer un lote de terreno dividido en parcelas. Si cada uno de los tratamientos se asigna aleatoriamente a una *ue* en el bloque, el arreglo obtenido se llama *diseño de bloques completos aleatorizados, DBA*.

El bloque es completo cuando todos los tratamientos aparecen en cada bloque. Cuando esta condición es clara, generalmente se suprime la palabra completo y se habla de diseño de bloques aleatorizados. Gráficamente, un DBA tiene la siguiente estructura (véase la figura 7.1).

Normalmente, en el DBA cada bloque es una repetición del experimento completo; así, el número de repeticiones es igual al número de bloques. En algunos experimentos un mismo bloque puede tener varias repeticiones de los tratamientos, con lo cual se obtiene un DBA generalizado o con repeticiones.

Puede ocurrir, además, que haya más tratamientos que *ue* por bloque, en cuyo caso los bloques quedan incompletos, pero aun así existen diseños, llamados de bloques incompletos, que permiten realizar el experimento satisfactoriamente.

	Tratamientos			
	T_1	T_2	- - -	T_k
Bloque 1			- - -	
Bloque 2			- - -	
⋮			⋮	
Bloque b			- - -	

Figura 7.1 Estructura del diseño de bloques

El primer problema en un DBA es cómo estratificar para que los bloques resulten homogéneos, pues la variabilidad de las *ue* tiene su origen en muchas fuentes; no obstante, algunas de estas pueden controlarse sistemáticamente. Para seleccionar aquella fuente de mayor variación, es prudente recurrir a la práctica y experiencia del investigador.

En todo experimento pueden reconocerse algunas variables que dificultan el control del error, puesto que influyen en el comportamiento de la variable de respuesta y no pueden controlarse físicamente en el experimento. Algunas de esas variables son intrínsecas a las *ue*, como el peso, la edad, el sexo y la procedencia genética o geográfica de los individuos; o externas, como los factores ambientales, el tiempo o el lugar donde se toman los datos.

Así, los bloques se construyen con unidades similares considerando la variable de mayor influencia y de modo que haya homogeneidad dentro de los bloques y heterogeneidad entre los mismos. De esta manera es posible aislar parte de la variabilidad de las *ue* y asignarla a la variable de bloqueo —variable o característica empleada para formar los bloques—, lo cual reduce sensiblemente el error experimental. Es claro que si la variación entre bloques es insignificante, no hay razón para realizar un DBA en vez de un DCA.

Además de la no disponibilidad de suficientes *ue* homogéneas para realizar un DCA, existe otro argumento en favor del DBA: este diseño permite expandir el espacio inferencial. Por ejemplo, cuando puede realizarse todo el experimento durante un solo día, un mes, etc., y este se repite en varios días, meses, etc., el tiempo puede considerarse como una variable de bloqueo. Así, la inferencia puede referirse a un periodo en el tiempo y no a un día específico.

Supóngase un diseño con k tratamientos, b bloques y una *ue* por tratamiento en cada bloque; este diseño requiere $n = kb$ *ue*, y la aleatorización de los tratamientos es únicamente dentro de las *ue* del bloque, por eso se dice que los bloques representan una restricción sobre la aleatorización del experimento.

7.2 Modelo matemático para un DBA y bloques fijos

En un diseño de bloques se contemplan dos clases de efectos: efectos de tratamientos, que son los de interés primario, y efectos de bloques, cuya magnitud debe aislarse del error experimental. Por otra parte, los datos de un experimento diseñado con bloques pueden organizarse en una tabla de doble entrada o de dos vías; es decir, pueden clasificarse de acuerdo con los niveles de dos factores: bloques y tratamientos. Aunque la variable, cuyo efecto se desea separar del error experimental, aparezca como el factor de bloqueo, no tiene la importancia de un factor de estudio en el experimento; de esta manera, el experimento continúa siendo unifactorial.

El modelo matemático para un DBA depende de la clasificación de los niveles del factor en estudio y de la forma de selección de los bloques. Los niveles del factor pueden ser fijos o aleatorios y los bloques pueden tomarse como un grupo fijo, aunque frecuentemente se seleccionan de un grupo mayor, o población de bloques, y se suponen aleatorios.

El modelo más simple corresponde a un diseño de bloques unifactorial con niveles fijos y un grupo de bloques igualmente fijo. Si se tiene un factor con k niveles y se construyen b bloques con una observación por celda, el modelo matemático tiene la forma:

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij} \text{ con } i = 1, 2, \ldots, b \text{ y } j = 1, 2, \ldots, k. \qquad (7.1)$$

Este modelo se diferencia del modelo (4.1), correspondiente a un DCA, en que tiene un sumando más que representa el efecto de bloques. Tanto el significado de los parámetros como los supuestos o restricciones en el modelo (4.1), son válidos para el modelo (7.1). Adicionalmente, respecto a los bloques se tiene:

1. β_i es el parámetro que representa el efecto del i-ésimo bloque, además $\Sigma_i\beta_i = 0$.
2. La variabilidad entre bloques es similar de un bloque a otro.
3. No hay efectos de *interacción* entre bloques y tratamientos.

El concepto estadístico de *interacción* se refiere a una relación entre combinaciones de tratamientos, la cual se verá en los diseños factoriales. Por el momento, la no interacción quiere decir que los efectos β_i y τ_j son aditivos y, por tanto, actúan independientemente en el experimento.

Para analizar el modelo (7.1) se emplea un análisis de varianza de dos vías (modelo I) cuya tabla correspondiente es la 7.1. En esta tabla, las sumas de cuadrados se calculan como en el DCA, teniendo en cuenta que existe una suma de cuadrados para bloques que se escribe como *SCB*.

La descomposición de la suma total de cuadrados es la siguiente:

$$\sum_i\sum_j\left(y_{ij}-\bar{y}_{..}\right)^2=\sum_i\sum_j\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2+\sum_i\sum_j\left(\bar{y}_{.j}-\bar{y}_{..}\right)^2+\sum_i\sum_j\left(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..}\right)^2,$$

y, en forma similar a la ecuación (4.3), se puede escribir que:

$$SCT = SCB + SCA + SCE.$$

Tabla 7.1 Tabla del ANOVA de dos vías y modelo I para un DBA

Fuente de variación	*SC*	*gl*	*MC*	*MC esparada*	*F*
Factor A	*SCA*	$k-1$	MCA	$\sigma^2 + \frac{b}{k-1}\sum \tau_j^2$	$\frac{MCA}{MCE}$
Bloques	*SCB*	$b-1$	MCB	$\sigma^2 + \frac{k}{b-1}\sum \beta_i^2$	
Error	*SCE*	$(b-1)(k-1)$	MCE	σ2	
Total	*SCT*	$kb-1$			

Ejemplo 7.1 Se diseñó un experimento para comparar la efectividad de ciertos compuestos químicos, como suplementos de la urea, para estimular el crecimiento de organismos celulolíticos de rumen de vacas Holstein, en periodo de lactancia. Las vacas se agruparon en grupos de tres, según la cantidad de leche producida durante las primeras cinco semanas de lactancia anteriores al tratamiento. Las raciones de maíz fueron suplementadas con: *a)* 80% de proteína de soya, *b)* urea y *c)* urea más químicos. Los datos de la tabla 7.2 son los registros de la producción promedio diaria durante 60 días (adaptado de Gill, 1978, vol. 2).

El primer paso para resolver este ejercicio consiste en chequear los supuestos del modelo matemático: la prueba de Kolmogorov-Smirnov para la normalidad de los residuos es no significativa ($p = 0.98$); las varianzas para raciones son homogéneas, como puede observarse en la tabla 7.2; la prueba de Burr-Foster, para igualdad de varianzas entre bloques, permite concluir homogeneidad ya que $q = 0.20 < 0.412 = q_{0.99,8,2}$, y, por último, el supuesto de aditividad o no interacción entre raciones y bloques es completamente válido, como se verá en la sección 7.4.

Usando calculadora, aunque lo recomendable es el computador, pueden obtenerse los siguientes resultados para construir la tabla del ANOVA.

$$C = \frac{1}{kb}\left(\sum_i \sum_j y_{ij}\right)^2 = (527.9)^2 / 8 \times 3 = 11611.6.$$

$$SCT = \sum_i \sum_j y_{ij}^2 - C = 12002.7 - 11611.6 = 391.1.$$

$$SCA = \frac{1}{b}\sum_j y_{.j}^2 - C = \left(\frac{1}{8}\right)\left(193.0^2 + 160.5^2 + 174.4^2\right) - 11611.6 = 66.5.$$

$$SCB = \frac{1}{k}\sum_j y_{i.}^2 - C = \left(\frac{1}{3}\right)\left(49.8^2 + 56.5^2 + \cdots + 83.6^2\right) - 11611.6.$$

$$SCE = SCT - SCA - SCB = 391.1 - 66.5 - 292.5 = 32.1.$$

Tabla 7.2 Producción promedio diaria de leche (kg)

Raciones	*Bloques (3 vacas)*								$y_{.j}$	s_j^2
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*		
a	17.3	22.6	22.1	21.7	24.7	28.8	25.9	29.9	193.0	16.80
b	14.5	17.2	19.0	18.1	19.2	20.8	24.2	27.5	160.5	16.83
c	18.0	16.7	19.7	21.7	21.6	24.5	26.0	26.2	174.4	12.74
$y_{i.}$	49.8	56.5	60.8	61.5	65.5	74.1	76.1	83.6	527.9	
s_i^2	3.43	10.7	2.64	4.32	7.60	16.03	1.02	3.52		

Tabla 7.3 Tabla del ANOVA para el ejemplo 7.1

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Raciones	66.5	2	33.2	14.5	<0.001
Bloques	292.5	7	41.8		
Error	32.1	14	2.3		
Total	391.1	23			

El factor de análisis es el alimento en sus diferentes raciones. Las hipótesis para comparar raciones son:

H_0: = $\mu_a = \mu_b = \mu_c$ contra H_1: no todas las medias son iguales.

La prueba F es altamente significativa (p = 0.0004), o sea, se puede concluir que la producción promedio de leche es diferente según el tipo de ración suministrada a la vaca.

Utilizando el método de Tukey para comparar raciones, se encuentra que $q_{0.05,3,14} = 3,7$, luego $DHS = 3.7\sqrt{2.3/8} = 1.98$, y los resultados son:

$a - b = 24.1 - 20.1 = 4.0 > 1.98$, diferencia significativa.

$a - c = 24.1 - 21.8 = 2.3 > 1.98$, diferencia significativa.

$c - b = 21.8 - 20.1 = 1.7 < 1.98$, diferencia no significativa.

Es interesante observar lo que hubiera ocurrido si el experimentador no hubiese tenido en cuenta la agrupación en bloques y hubiese considerado un DCA que, por azar, resultara igual al de la tabla 7.2. El ANOVA correspondiente a este diseño es el de la tabla 7.4. Dos aspectos de esta tabla son relevantes. En primer lugar, la suma de cuadrados del error es 324.6, la cual rebajó a 32.1 en el DBA. En este caso la disminución fue sustancial, aunque por otro lado, los *gl* del error pasaron de 21 a 14 en el DBA. En conclusión, no todo fue ganancia al diseñar en bloques: disminuyó la suma de cuadrados del error pero se perdieron varios *gl*. Sin embargo, la *MCE* pasó de 15.5 a 2.3, con lo cual se logra mayor potencia en el DBA.

El segundo aspecto tiene que ver con la prueba F. La producción de leche en el periodo inicial es una variable intrínseca al animal que debe controlarse para evitar un error experimental alto. Esta variable fue tan influyente en la variable de respuesta que no permitió diferenciar la eficacia de las raciones (p = 0.14).

Para que el diseño de bloques sea eficiente, la reducción en el error experimental debe contrarrestar la pérdida correspondiente de *gl* de la suma de cuadrados del error. Antes de elegir un DBA debe tenerse la seguridad de que los bloques queden formados por unidades verdaderamente homogéneas, en caso contrario, podría ser más conveniente el DCA.

De acuerdo con Montgomery (2007), si existen dudas sobre la importancia de considerar bloques en un experimento, deberá optarse por el diseño de bloques esperando que el efecto de estos sea grande. Si este en realidad no existe, y el

Tabla 7.4 Tabla del ANOVA para el ejemplo 7.1 considerado incorrectamente como un DCA

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Raciones	66.5	2	33.2	2.15	0.14
Error	324.6	21	15.5		
Total	391.1	23			

número de *gl* del error es moderadamente grande, no hay que preocuparse por cambiar de diseño.

La eficiencia de un diseño de bloques con respecto a un DCA equivalente, puede calcularse comparando sus variabilidades. La interpretación respectiva se verá en la siguiente sección.

7.3 Restricción de aleatoriedad y eficiencia de un DBA

El investigador podría estar interesado en probar diferencias entre medias de bloques, por ejemplo, usando el cociente $F = MCB / MCE$. Sin embargo, deberá tener cuidado en la interpretación de esta prueba, ya que los bloques fueron aleatorizados internamente pero no externamente. Esto es, los *k* tratamientos fueron aleatorizados para las *k ue* dentro del bloque, pero las *kb* combinaciones de tratamientos no se asignaron al azar a las respectivas *ue*. Bajo estas circunstancias, la prueba *F* para la hipótesis H_0: $\beta_i = 0$ es inapropiada porque los bloques fueron diseñados en forma no aleatoria, es decir, la aleatorización fue restringida.

Anderson y McLean (1974) presentan un análisis detallado de este problema. Según ellos, hay una restricción en la construcción del diseño con respecto a la aleatorización en el sentido de que la única aleatorización es la que se realiza separadamente dentro de cada bloque. Esta restricción debería aparecer en el modelo matemático y es, a la vez, otra fuente de variación, pero como no es estimable de los datos, o sea, está confundida con el efecto de bloques, generalmente no aparece en la tabla del ANOVA. Lo más importante es que el investigador piense en la restricción que le impuso a su diseño y entienda cuál es su efecto en el resultado final del experimento.

Ahora bien, la prueba *F* para bloques es válida cuando el interés no es comparar medias sino constatar la efectividad del diseño en la reducción del error experimental. El cociente $F = 41.8 / 2.3 = 18.2$ indica que la prueba es altamente significativa ($p < 0.001$) y, por tanto, que se tuvo éxito al diseñar los bloques. Por otro lado, la conclusión respecto a la diferencia significativa entre medias de bloques no es válida; pero esto no debe preocupar, ya que los bloques son una fuente de variación reconocida de experimentos previos y se espera que las diferencias resulten significativas.

Más importante que el resultado de la prueba anterior, es el cálculo de la eficiencia relativa del diseño. Un resultado significativo corrobora que la formación de bloques fue efectiva; entonces, lo que sigue es cuantificar esa eficiencia.

La eficiencia relativa del DBA con respecto a un DCA equivalente, esto es, que tiene los mismos datos, es:

$$E_R = \left[\frac{(b-1)MCB + b(k-1)MCE}{(kb-1)MCE}\right]\left[\frac{(\upsilon_1+1)(\upsilon_2+3)}{(\upsilon_1+3)(\upsilon_2+1)}\right],$$

donde *MCB* es la media cuadrática de bloques, *MCE* es la media cuadrática del error en el DBA, υ_1 es el número de *gl* del error en el DBA e υ_2 es el número de *gl* del error en el DCA.

Aplicada a los datos del ejemplo 7.1, resulta:

$$E_R = \left[\frac{(7\times 41.8)(8\times 2\times 2.3)}{23\times 2.3}\right]\left[\frac{(14+1)(21+3)}{(14+3)(21+1)}\right] = 6.02.$$

Si se hubiese utilizado un DCA, sería necesario multiplicar por seis el número de repeticiones para obtener la misma precisión conseguida con el DBA, esto es, un total de $6.02 \times 8 = 48$ repeticiones, en vez de ocho, para obtener el mismo error estándar de las medias de los tratamientos. En términos porcentuales, el DBA tuvo una eficiencia del 502%.

7.4 Supuestos del modelo y prueba de aditividad

El examen crítico, sea de los datos crudos o de los residuos en un DBA, sigue de cerca la metodología discutida en el capítulo 5. Es necesario, además, realizar un chequeo sobre la aditividad entre bloques y tratamientos. Aunque los bloques representan influencias que deben eliminarse en vez de estudiarse, pueden presentarse efectos inesperados importantes que solo se reconocen en un análisis cuidadoso de los residuos; este es el caso de las interacciones entre bloques y el factor de análisis.

El análisis gráfico incluye residuos de bloques y residuos de tratamientos contra observaciones. Un gráfico que merece especial atención es el de residuos contra valores ajustados (e_{ij} contra $\hat{y}_{ij}$) como el de la figura 7.2. Si el modelo es adecuado, este gráfico no muestra asociación entre los tamaños de e_{ij} y de $\hat{y}_{ij}$, por el contrario, si el gráfico presenta una forma de embudo similar a la de la figura 5.7, es porque hay asociación entre algunos parámetros del modelo. Una forma curvilínea —generalmente cuadrática— del gráfico indica no aditividad entre efectos de tratamientos y de bloques. Detectada la no aditividad gráficamente, puede aplicarse la prueba de aditividad de Tukey (1949); si esta resulta significativa, puede pensarse en una transformación de los datos para eliminar la interacción. No obstante, algunos autores recomiendan rediseñar el experimento considerando los bloques como otro factor y repitiendo observaciones por celda; esto para no perder la información dada por las interacciones.

La no aditividad en un DBA es quizá la falla más crítica, puesto que su consecuencia es aumentar el error experimental, lo cual implica inexactitudes en estimaciones y pruebas.

La prueba de Tukey es un chequeo a posteriori del supuesto de aditividad (véase tabla 7.5). Aunque la figura 7.2 sugiere aditividad en los datos que se han estado manejando, es conveniente realizar la prueba con el fin de indicar el procedimiento que se seguiría en un caso diferente.

Las hipótesis que se deben probar son de la forma H_0: β_i y τ_j son aditivos contra H_1: hay interacción entre β_i y τ_j. El procedimiento consiste en subdividir la *SCE* en dos componentes —una de ellas llamada suma de cuadrados de no aditividad con 1 *gl*— y compararlas mediante una prueba *F*. La suma de cuadrados de no aditividad se calcula como: *SC* (no aditividad) $= w^2 / D$, y se obtiene como sigue:

$$D = \left(\sum_i \hat{\beta}_i^2\right)\left(\sum_j \hat{\tau}_j^2\right), \text{ donde } \hat{\beta}_i = \bar{y}_{i.} - \bar{y}_{..} \text{ y } \hat{\tau}_j = \bar{y}_{.j} - \bar{y}_{..}.$$

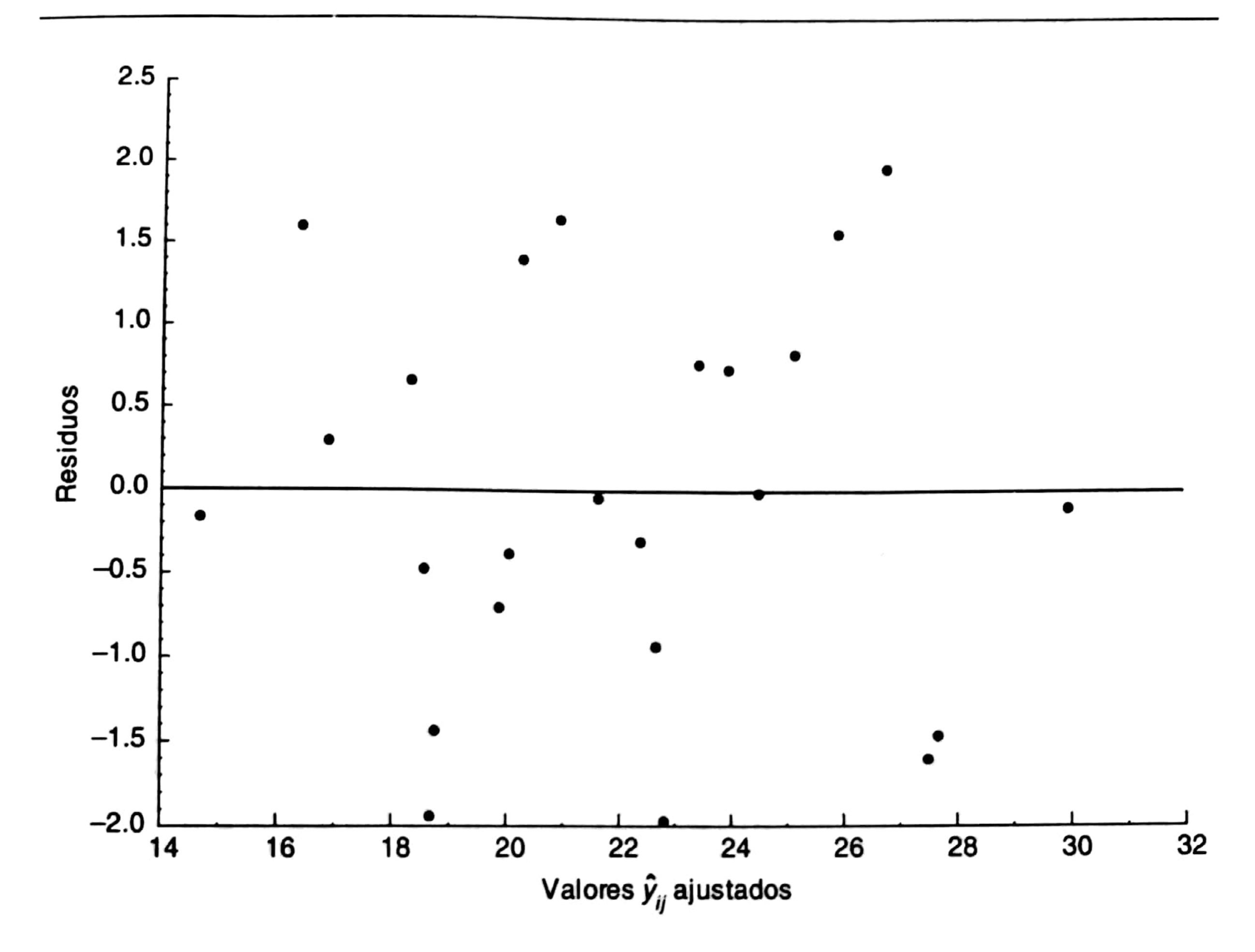


Figura 7.2 Gráfico de residuos contra valores ajustados $\bar{y}_{ij}$

Tabla 7.5 Datos para la prueba de aditividad de Tukey

Raciones	*Bloques* 1	2	3	4	5	6	7	8	$\bar{y}_{.j}$	$\hat{\tau}_j$
a	17.3	22.6	22.1	21.7	24.7	28.8	25.9	29.9	24.1	2.1
b	14.5	17.2	19.0	18.1	19.2	20.8	24.2	27.5	20.1	–1.9
c	18.0	16.7	19.7	21.7	21.6	24.5	26.0	26.2	21.8	–0.2
$\bar{y}_{i.}$	16.6	18.8	20.3	20.5	21.8	24.7	25.4	27.9	22.0	
$\hat{\beta}_i$	–5.4	–3.2	–1.7	–1.5	–0.2	2.7	3.4	5.9		
W_i	5.18	11.44	6.37	6.84	11.07	16.06	3.21	5.30		

Por ejemplo:

$\hat{\beta}_1 = 16.6 - 22.0 = -5.4$, $\hat{\beta}_2 = 18.8 - 22.0 = -3.2$, etc.

Entonces $D = (98.24)(8.06) = 791.8$.

Por otra parte, $w_i = \sum_j y_{ij}\tau_j$, así:

$w_1 = 17.3(2.1) + 14.5(-1.9) + 18.0(-0.2) = 5.18,$

$w_2 = 22.6(2.1) + 17.2(-1.9) + 16.7(-0.2) = 11.44$, etc.

Finalmente, $w = \sum_i w_i \hat{\beta}_i = -2.337$, y $w^2 = 5.46$.

Por tanto, SC (no aditividad) = 5.46 / 791.8 = 0.007.

La tabla 7.6 muestra la información para la prueba de aditividad según Tukey. Se concluye que la no aditividad es no significativa, o más claramente, que el supuesto de aditividad es válido. En esta prueba no se debería seleccionar un α muy pequeño, porque la prueba no sería sensible. Gill (1978, vol. 2) recomienda $\alpha = 0.25$ como valor adecuado.

7.5 Bloques aleatorios y modelos mixtos

Los diseños de bloques son fácilmente adaptables a gran variedad de situaciones. En el ejemplo 7.1, tanto el factor de estudio como el de bloques fueron considerados de efectos fijos, pero es más común el diseño de bloques aleatorios, es decir, aquel donde los bloques seleccionados son representativos de una población grande de bloques. Este diseño puede dar origen a un modelo aleatorio si el factor es de efectos aleatorios, o a un modelo mixto si el factor es de efectos fijos.

El ejemplo 7.2 se analiza mediante un ANOVA para el modelo mixto y, aunque este análisis es el mismo cuando se trata del modelo aleatorio, debe tenerse en cuenta que hay diferencias en la interpretación de los resultados. El modelo matemático tiene las formas:

1. $y_{ij} = \mu + B_i + \tau_j + \varepsilon_{ij}$; modelo mixto, (7.2)

donde las B_i son variables aleatorias independientes con distribución $N(0,\sigma_B^2)$.

2. $y_{ij} = \mu + B_i + A_j + \varepsilon_{ij}$; modelo aleatorio, (7.3)

con la misma interpretación del modelo (4.4).

En la tabla del ANOVA para estos modelos (véase tabla 7.7), se observa que la prueba F para tratamientos —factor A— puede establecerse mediante el cociente MCA / MCE en ambos casos.

Tabla 7.6 Tabla del ANOVA para una DBA, incluyendo la no aditividad como otra fuente de variación

Fuente de variación	*SC*	*gl*	*MC*	*F*
Raciones	66.5	2	33.250	13.467
Bloques	292.5	7		
Error	32.1	14		
No aditividad	0.007	1	0.007	0.003 ns
Residuo	32.093	13	2.469	

También puede diseñarse el modelo mixto con bloques fijos y factor aleatorio, pero este sería equivalente al modelo (4.4) ya analizado.

Ejemplo 7.2 Para comparar el efecto de cinco dietas en la ganancia de peso en ratas, durante las cuatro semanas siguientes al destete, se planeó un diseño de bloques tomando camadas de cinco ratas. Cada camada constituyó un bloque, de esta forma se trató de eliminar la influencia de la variabilidad genética en la ganancia de peso. Los datos se dan en la tabla 7.8 (adaptado de John y Quenouille, 1977).

Las sumas de cuadrados para construir la tabla del ANOVA se calcularon de la misma forma que en el ejemplo 7.1, con lo cual se obtuvo la tabla 7.9.

El factor que se está analizando es la dieta con cinco niveles. Las hipótesis son:

H_0: $\mu_A = \mu_B = \mu_C = \mu_D = \mu_E$.

H_1: no todas las medias son iguales.

La prueba F es no significativa, luego no hay suficiente evidencia para reconocer diferencias en la ganancia de peso causada por las dietas y, por tanto, no se hacen comparaciones por pares entre las medias.

Independientemente de este resultado, el investigador puede probar comparaciones diseñadas desde el inicio del experimento. Como ejemplo, supóngase que la dieta E, rica en harina de pescado, se incluyó para compararla con las cuatro primeras. El contraste es de la forma $Q = (\mu_1 + \mu_2 + \mu_3 + \mu_4) / 4 - \mu_5$. Su significancia se prueba calculando:

$\hat{Q} = (67.8 + 70.2 + 68.0 + 69.8)/4 - 78.5 = -9.6.$

$SC(\hat{Q}) = b\hat{Q}^2 / \sum c_j^2 = 6(9.6)^2 / (5/4) = 442.4.$

$F = SC(\hat{Q}) / MCE = 442.4/41.7 = 10.6$. Este resultado es altamente significativo ($p < 0.005$).

El chequeo de los supuestos para la validez del ANOVA dio estos resultados: la prueba de normalidad de Kolmogorov-Smirnov es no significativa ($p = 0.31$); la prueba de Bartlett permite concluir que no hay heterogeneidad de varianzas, tanto para dietas ($p = 0.84$) como para camadas ($p = 0.96$); en el modelo de bloques aleatorios no es requisito la aditividad, como se verá a continuación; y, por último, en este ejemplo es suficiente con mirar los gráficos de residuos para concluir que los supuestos se satisfacen plenamente.

La eficiencia del diseño es $E_R = 2.51$, esto quiere decir que se requieren $2.51 \times 6 = 15$ repeticiones para obtener igual precisión con un DCA equivalente.

El supuesto de aditividad es requerido en el modelo fijo mas no en el modelo aleatorio o en el mixto. Si hay interacción, el modelo será:

Tabla 7.7 Tabla del ANOVA para bloques aleatorios y una repetición por celda

				MC esperada	
Fuente de variación	*SC*	*gl*	*MC*	*Modelo (7.2)*	*Modelo (7.3)*
Factor A	SCA	$k-1$	*MCA*	$\sigma^2 + \frac{b}{k-1}\sum \tau_j^2$	$\sigma^2 + b\sigma_A^2$
Bloques	SCB	$b-1$	*MCB*	$\sigma^2 + k\sigma_B^2$	$\sigma^2 + k\sigma_B^2$
Error	SCE	$(b-1)(k-1)$	*MCE*	σ^2	σ^2
Total	SCT	$kb-1$			

$$y_{ij} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \varepsilon_{ij}; \text{ modelo fijo,} \tag{7.4}$$

donde $(\beta\tau)_{ij}$ es el efecto fijo de la interacción del bloque i con el tratamiento j. El ANOVA correspondiente se da en la tabla 7.10; allí se observa que la media cuadrática del error no es estimable —tiene 0 *gl*—, por lo que no puede efectuarse la prueba con $F = MCA / MCE$. Sin embargo, se puede efectuar la prueba para diferencia entre tratamientos mediante $F = MCA / MCAB$, donde $MCAB$ es la media cuadrática de la interacción. Pero dicha prueba resulta muy conservadora ya que el denominador puede ser mayor que su valor esperado. Además, la prueba puede resultar no significativa cuando en realidad los tratamientos difieren entre sí. Entonces, si la prueba resulta significativa, puede aceptarse sin problemas; pero si no es significativa, es conveniente probar, en el diseño original, la no aditividad o repetir observaciones en celdas para estimar la *MCE*.

Cuando los bloques son aleatorios, el modelo tiene la forma:

$$y_{ij} = \mu + B_i + \tau_j + (B\tau)_{ij} + \varepsilon_{ij}; \text{ modelo mixto,} \tag{7.5}$$

donde $(B\tau)_{ij}$ es el efecto mixto de interacción de bloques y tratamientos y $\sigma^2_{B\tau}$ es la componente mixta de la variabilidad de cada observación. El ANOVA para este modelo se muestra en la tabla 7.10.

En el modelo mixto, la prueba para el factor A es $F = MCA / MCAB$, la cual es válida sin importar que la interacción sea o no significativa. Para este modelo, el error

Tabla 7.8 Ganancia de peso (en g) de ratas sometidas a cinco dietas diferentes

			Dietas				
Camadas	*A*	*B*	*C*	*D*	*E*	$y_{i.}$	s_i^2
1	57	65	70	68	76	336	48.7
2	55	66	59	67	75	322	59.8
3	62	69	64	69	77	341	33.7
4	75	61	74	72	87	369	85.7
5	86	91	78	90	95	440	41.5
6	72	69	63	53	61	318	54.8
$y_{.j}$	407	421	408	419	471	2126	
$\bar{y}_{.j}$	67.8	70.2	68.0	69.8	78.5	70.9	
s_j^2	143.0	113.0	52.4	141.4	134.3		

Tabla 7.9 Tabla del ANOVA para el ejemplo 7.2

Fuente de variación	*SC*	*gl*	*MC*	F	*p*
Dietas	464	4	116.0	2.78	0.055
Camadas	2087	5	417.4		
Error	833	20	41.7		
Total	3384	29			

estándar de la media de tratamientos no es estimable, pero el error estándar de la diferencia de medias sí puede estimarse de los datos, Anderson y McLean (1974).

Por otra parte, en todo diseño de bloques es posible tener más de una repetición por cada tratamiento en cada bloque. La repetición puede ser otra medición de la misma *ue* o una medición en otra *ue*. Si cada tratamiento se repite *s* veces en cada bloque, se obtiene un diseño de bloques aleatorizados generalizado. Hay que tener mucho cuidado en lo que se entiende por repetición, pues cada interpretación da lugar a un modelo matemático diferente y, por tanto, a un ANOVA también diferente.

Considérese el caso en que cada *ue* origina dos o más lecturas para el mismo tratamiento. Por ejemplo, si preparar una *ue* resulta costoso y la observación hecha tiene un error de medida grande pero puede hacerse con prontitud y bajo costo, resulta económico y eficiente hacer varias observaciones o medidas sobre la misma *ue* para reducir los efectos de los errores de medida. En esta situación, las repeticiones no tienen mucho valor en cuanto a la ampliación del campo de la inferencia.

Ahora considérese un experimento de invernadero; las *ue* pueden ser las materas y las localizaciones conformarían los bloques. Si se desea precisión en las lecturas, estas pueden repetirse dividiendo el material de cada matera. Una situación diferente se da si cada tratamiento se asigna al azar a varias materas en el mismo bloque, y se obtiene una sola lectura de cada matera. La inferencia es más general en este caso que en el anterior.

Otro experimento diferente de los anteriores, consiste en submuestrar cada *ue*. Es decir, si la *ue* es una matera con varias plantas, a veces es difícil medir toda la unidad experimental, por lo que se recurre al submuestreo aceptando que se introduce otra fuente de variabilidad en el experimento.

En conclusión, cada modalidad plantea un modelo matemático diferente y le agrega cierto grado de dificultad a los análisis de varianza anteriores. Estos modelos se discuten en Anderson y McLean (1974), en Gill (1978, vol. 2) y en Steel y Torrie (1997).

Tabla 7.10 Tabla del ANOVA para un DBA con interacción y una repetición por celda

Fuente de variación	*gl*	*MC*	*MC esperada* Modelo (7.4)	*MC esperada* Modelo (7.5)
Factor A	$k-1$	MCA	$\sigma^2 + \frac{b}{k-1}\sum\tau_j^2$	$\sigma^2+\sigma_{B\tau}^2 + \frac{k}{b-1}\sum\tau_j^2$
Bloques	$b-1$	MCB	$\sigma^2 + \frac{k}{b-1}\sum\beta_i^2$	$\sigma^2 + k\sigma_B^2$
Interacción	$(b-1)\,(k-1)$	MCAB	$\sigma^2 + \frac{\left[\sum\sum(\tau\beta)_{ij}^2\right]}{(b-1)\; k-1}$	$\sigma^2 + \sigma_{B\tau}^2$
Error	0	MCE	σ^2	σ^2
Total	$kb-1$			

7.6 Datos perdidos y diseños no ortogonales

En los diseños multifactoriales, la ortogonalidad del diseño significa que los estimadores de los parámetros asociados con un factor no están correlacionados con los de los otros factores. En particular, si cada nivel de un factor aparece el mismo número de veces con los niveles del otro factor, el diseño será ortogonal para ambos factores. Si uno de los factores es una variable de bloqueo, es suficiente con que todos los tratamientos aparezcan una sola vez en cada bloque para que el DBA sea ortogonal. Si falta una o más observaciones en los bloques, el diseño deja de ser ortogonal.

Los diseños ortogonales tienen la ventaja de la simplicidad en el ANOVA correspondiente: las sumas de cuadrados se calculan de fórmulas establecidas para cada caso; mientras que en diseños no ortogonales las sumas de cuadrados no pueden estandarizarse, por consiguiente, su análisis resulta más difícil.

El problema de uno o más datos faltantes para el análisis final no presenta inconvenientes en un DCA pero sí en un DBA, porque en este se destruye la ortogonalidad. Un dato puede perderse porque la *ue* desapareció; el individuo enfermó, murió o abandonó el tratamiento; el cultivo en la parcela fue destruido, o la semilla no germinó; un vial se destruyó, el aparato medidor se dañó, la persona encargada del control se descuidó, etc.; es decir, porque la medición no pudo hacerse por una u otra razón y, por tanto, la información resulta irrecuperable.

Existen dos métodos para resolver el problema de una o más celdas vacías. Uno de ellos consiste en aplicar las técnicas de la regresión; este procedimiento puede ser complejo pero proporciona una solución exacta al problema, como se muestra en Hicks y Turner (1999). Otro procedimiento, más sencillo pero aproximado, consiste en estimar el valor faltante minimizando la suma de cuadrados del error.

Para un solo valor faltante y_{ij}, este último método da la siguiente fórmula de estimación:

$$\hat{y}_{ij} = \frac{by'_{i.} + ky'_{.j} - y'_{..}}{(b-1)(k-1)},$$

donde $y'_{i.}$, es el total del bloque que tiene la celda vacía, $y'_{.j}$ es el total del tratamiento de la celda vacía y $y'_{..}$ es el gran total, el cual no incluye la celda vacía.

Ejemplo 7.3 Realizar el ANOVA para el ejemplo 7.1 suponiendo que la celda y_{41} está vacía —el subíndice i para bloques y el subíndice j para raciones.

Se suprime el valor $y_{41} = 21.7$, en la tabla 7.2, y se recalculan las sumas siguientes:

$y_{4.}' = 61.5 - 21.7 = 39.8$, suma del bloque,

$y_{.1}' = 193.0 - 21.7 = 171.3$, suma del tratamiento,

$y_{..}' = 527.9 - 21.7 = 506.2$, suma de las observaciones.

Entonces,

$$\hat{y}_{41} = \frac{(8 \times 39.8) + (3 \times 171.3) - (506.2)}{7 \times 2} = 23.3.$$

Se llena la celda vacía con el número 23.3 y se realiza nuevamente el ANOVA restándole 1 a los *gl* del error. Este enfoque, que es aproximado, genera un sesgo en la prueba *F*,

lo cual conduce a rechazar la hipótesis H_0 con frecuencia mayor que la esperada, esto es, el valor p obtenido es menor que el verdadero. Para corregir esta desviación, se disminuye la *SCA* en una cantidad B que estima el sesgo así:

$B = \frac{\left[y_{i.}' - (k-1)\hat{y}_{ij}\right]^2}{k(k-1)}$. Para los datos anteriores:

$B = (1/6)(39.8 - 2 \times 23.3)^2 = 7.7$, de modo que $SCA = 73.5 - 7.7 = 65.8$.

Los resultados se dan en la tabla 7.11; se observa que son similares a los de la tabla 7.3.

El error estándar de la diferencia entre la media del tratamiento que tiene la celda vacía y la media de otro tratamiento cualquiera, se calcula con la fórmula:

$EE = \{MCE\,[2/b + k/b(b-1)(k-1)]\}^{1/2}$.

Cuando hay dos o más datos perdidos, las estimaciones se calculan mediante un proceso iterativo, como lo explican Li (1969a) y Cochran y Cox (1995).

En algunas investigaciones puede ser interesante estimar el número de bloques necesarios para alcanzar una potencia determinada. Al aumentar los bloques aumenta el número de *gl* del error; por tanto, si el modelo es de efectos fijos, puede usarse el mismo procedimiento descrito en la sección 4.6 con $\phi^2 = b\Sigma\tau_j^2 / k\sigma^2$ y, si el modelo es aleatorio, con $\lambda^2 = 1 + b\sigma_A^2 / \sigma^2$ teniendo en cuenta que $\nu_1 = k - 1$ y $\nu_2 = (b-1)(k-1)$.

7.7 Diseños de bloques con medidas repetidas

El ANOVA para un diseño de bloques puede considerarse como una generalización de la prueba t en el diseño de muestras pareadas. Los datos pareados se obtienen del mismo sujeto en dos tiempos diferentes, por ejemplo antes y después de la aplicación de un tratamiento. Por otra parte, un bloque puede considerarse como la generalización de un par; por ejemplo, cuando el sujeto hace las veces de bloque al recibir tres o más tratamientos en secuencia, o al ser observado en tres o más tiempos o lugares diferentes.

Cuando el sujeto, o en general la *ue* sirve de bloque, no puede hablarse de unidades experimentales dentro del bloque, sino que estas deben tomarse como los diferentes momentos en los que el sujeto recibe los tratamientos. En este caso, se aleatoriza individualmente el orden de aplicación de los tratamientos a cada

Tabla 7.11 Tabla del ANOVA aproximado para el ejemplo 7.1 con un dato faltante

Fuente de variación	*SC*	*gl*	*MC*	*F*
Raciones	65.8	2	32.9	14.0
Bloques	288.5	7	41.2	
Error	30.6	13	2.3	
Total	392.6	22		

sujeto. El diseño anterior se conoce como DBA con medidas repetidas o diseño de tratamientos por sujetos.

Dos ventajas que se pueden resaltar en este tipo de diseño son la posibilidad de aislar del error experimental todas aquellas fuentes de variabilidad entre sujetos —cada sujeto sirve como su propio control—, y el número reducido de *ue* necesarias: aspectos que son particularmente importantes cuando solo se tienen unas pocas *ue* disponibles para el experimento.

Muchas veces la unidad experimental proporciona las respuestas en un orden fijo en el tiempo o en el espacio; por ejemplo, a diferentes horas del día o épocas del año, a diferentes profundidades en un sitio de un lago, a diferentes alturas en un árbol o en diferentes partes del cuerpo. Aquí el término tratamiento es únicamente un nombre adecuado para las medidas repetidas realizadas secuencialmente o en un orden preestablecido.

Los diseños de medidas repetidas pueden originar correlaciones intrabloque. Por ejemplo, algunos tratamientos producen efectos residuales que influyen en el resultado de los tratamientos siguientes o, como en muchos animales —vacas lecheras, gallinas ponedoras, etc.—, la producción declina con el tiempo, lo cual origina tendencias que deben tenerse en cuenta en la aplicación de los tratamientos.

Para corregir lo anterior, se utiliza la aleatorización de los tratamientos, aunque no siempre puede aplicarse. También puede recurrirse a diferentes variaciones en los diseños; una de ellas es dejar que pase un tiempo suficiente entre las aplicaciones de un tratamiento y el siguiente para minimizar el posible efecto residual, llamado efecto de arrastre *(carryover)*. A veces es conveniente balancear el orden de aplicación de los tratamientos y entonces se usan otros diseños, como los conmutativos —*crossover*— o los de cuadrados latinos.

Gill (1978, vol. 2) y Arnau (1981, vol. 1) analizan diseños de medidas repetidas, usando modelos que incluyen la correlación entre periodos o tratamientos para medir el efecto residual y aislarlo del error experimental.

En los DBA con medidas repetidas, los sujetos se consideran como una muestra seleccionada al azar de una población, así que el modelo matemático puede ser uno de los tipos (7.2), (7.3) o (7.5), suponiendo que no existen efectos residuales ni tendencias específicas en los periodos de tratamiento. Además el efecto B_i no representa la variabilidad entre sujetos de un mismo bloque sino la variabilidad de cada sujeto individual bajo diferentes tratamientos.

Algunos paquetes estadísticos enfocan el análisis de varianza de factores de medidas repetidas no como un diseño univariado sino multivariado. En este caso, el análisis de varianza multivariado es menos exigente en cuanto a los supuestos del modelo.

En el ejemplo siguiente, cada planta se considera como un bloque y se supone que no hay correlación entre las lecturas en las diferentes horas del día.

Ejemplo 7.4 En un estudio bioecológico en *Guzmania sp.* (Bromeliaceae) y *Heliconia bihai* (Musaceae) en un bosque pluvial del departamento de Chocó (Colombia), Paz (1980) midió, a intervalos de seis horas durante el día, la concentración de oxígeno en el contenido líquido de las brácteas de las plantas. Interesa conocer si la concentración de oxígeno cambia con las horas del día (véase tabla 7.12).

De acuerdo con el modelo (7.2) y la tabla 7.7, se realizó el ANOVA que se muestra

en la tabla 7.13. En cuanto a los supuestos, chequeados en el computador, puede concluirse que todos son válidos, incluyendo el de aditividad. Como las plantas, en este caso los bloques, son una muestra de las plantas del bosque, y las horas fueron escogidas a intervalos fijos, el modelo apropiado para este estudio es el modelo mixto. Aun si hubiese interacción entre plantas y horas del día (modelo 7.5), el ANOVA hubiese sido el mismo y la conclusión seguiría siendo que no hay diferencias en la concentración de oxígeno durante el día ($p = 0.84$).

Tabla 7.12 Concentración de O_2 (en mg/L) en el contenido líquido de las brácteas de plantas durante el día, según el ejemplo 7.4

Planta	*6:00 h*	*12:00 h*	*18:00 h*	*00:00 h*
1	4.1	4.6	5.7	4.2
2	2.5	1.0	1.5	3.0
3	3.7	3.5	5.6	3.9
4	4.0	5.2	5.4	4.1
5	3.0	4.0	3.8	3.5
6	4.2	6.2	5.4	4.6
7	2.4	1.6	1.3	3.3
8	6.6	4.5	3.6	2.7
9	2.6	2.7	3.1	2.5
10	3.8	2.9	5.8	3.2
11	4.2	4.4	5.0	3.9
12	2.4	2.3	2.9	2.5
13	5.6	5.2	5.1	7.4
14	1.6	3.7	1.8	3.7
15	4.8	2.1	4.1	6.4
16	2.4	2.8	2.5	2.9
17	5.5	3.4	4.6	4.8
18	5.1	3.1	5.4	5.7
19	3.2	3.1	3.4	2.6
20	2.4	4.9	3.5	2.9
21	3.1	3.4	3.2	2.6
22	2.5	5.3	3.9	3.0
23	3.6	4.2	2.6	3.8
24	3.5	3.4	2.8	3.2
Totales	86.8	87.5	92.0	90.4
Varianzas	1.54	1.59	1.91	1.59

Tabla 7.13 Tabla del ANOVA para el ejemplo 7.4

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Horas	0.75	3	0.25	0.28	0.84
Plantas	91.11	23	3.96	4.44	
Error	61.62	69	0.89		
Total	153.48	95			

Con el modelo (7.2) se puede probar la variabilidad entre plantas; se encuentra que esta es altamente significativa ($p = 0.001$). Si el modelo para emplear fuera el (7.5), esta prueba puede resultar incorrecta.

El número de plantas requeridas en una investigación similar, suponiendo una potencia no inferior a 0.90, se estima con los métodos de la sección 4.6. Para $\alpha = 0.05$, $\nu_1 = 3$, $\nu_2 = 69$, $\sigma^2 = 0.89$, $\tau_j = 0.5$ —valor arbitrario pero basado en las concentraciones de O_2 entre una hora y otra—, y $\Sigma\tau_j^2 = 1.0$, se tiene que $\phi^2 = (24 \times 1.0) / (4 \times 0.89) = 6.74$ y $\phi = 2.6$. Mirando la curva correspondiente a $\nu_1 = 3$, $\alpha = 0.05$ y $\phi = 2.6$, se encuentra que $\beta = 0.0$. Se concluye que 24 bloques es un número bastante grande, entonces se ensaya con valores más pequeños y se encuentra que con $b = 14$, $\phi = 1.98$ y se lee un error $\beta = 0.09$. Bajo las condiciones anteriores, basta con analizar catorce plantas para obtener una potencia de alrededor de 0.90 en la prueba F del ANOVA.

7.8 Prueba de rangos de Friedman para un DBA

El análisis paramétrico del diseño de bloques es válido cuando se cumplen todos los supuestos para los modelos (7.1), (7.2) y (7.3). Las observaciones en los diferentes bloques son independientes; pero entre tratamientos, estas pueden presentar asociaciones que invalidan las conclusiones del ANOVA.

Friedman (1937) propuso reemplazar las observaciones en cada bloque por sus respectivos rangos R_{ij}, tomados en el orden de magnitud de cada observación con respecto a las demás del mismo bloque, cuando los supuestos de independencia, normalidad u homogeneidad no son sustentables. De otro lado, bajo el supuesto de bloques independientes y datos provenientes de una población continua, puede aplicarse la prueba F a los rangos en vez de a los datos originales.

Las propiedades de los rangos hacen que no sea necesario calcular directamente las sumas de cuadrados del ANOVA. La prueba sobre diferencias entre tratamientos puede realizarse en una forma más sencilla, como sigue:

1. Se asignan rangos de menor a mayor a las k observaciones de cada bloque.
2. Se obtiene la suma de los rangos para cada tratamiento así:

$$R_{.j} = \sum_i R_{ij}.$$

3. Se calcula la estadística $S = \frac{12}{kb(k+1)} \sum R_{.j}^2 - 3b(k+1)$.

4. Para un nivel de significación α, se rechaza H_0 si $S \geq \chi^2_{\alpha, k-1}$. Para valores pequeños de b y de k la prueba es aproximada, por consiguiente, deberán usarse tablas especiales en vez de la distribución χ^2 (Hollander y Wolfe, 1999). Igualmente, si aparecen varios empates, es conveniente calcular un valor de corrección para S, como se hará en el ejemplo siguiente.

Ejemplo 7.5 Se realizó un experimento para investigar el efecto tóxico de tres sustancias químicas, A, B y C, sobre la piel de las ratas. Para ello se tomaron ocho ratas y en el lomo de cada una se marcaron tres secciones cuadradas y adyacentes de 2.5 cm de lado. Luego se aplicaron las tres sustancias y se anotó una puntuación de 0 a 10 dependiendo del grado de irritación de la piel. Cada rata se consideró como un bloque. Los resultados aparecen listados en la tabla 7.14 (adaptado de Mendenhall, 1968).

Se aplica la prueba de Friedman para comparar la toxicidad de las tres sustancias. Los rangos de los datos aparecen entre paréntesis en la misma tabla y en la parte inferior las sumas de los rangos.

Se calcula la estadística S así:

$$S=\left[\frac{12}{(8\times3\times4)}\right]\left[(14.0)^2+(21.5)^2+(12.5)^2\right]-(3\times8\times4)=5.81.$$

Hay observaciones empatadas en los bloques 2, 5, 6 y 8. La corrección para S consiste en dividir su valor por el factor C, que se calcula de la siguiente manera:

$$C=1-\frac{\sum_i T_i}{kb\left(k^2-1\right)};\ \text{con}\ T_i=\sum_h t_{ih}^3-\sum_h t_{ih}.$$

donde t_{ih}, es el número de datos empatados para un rango dado en el bloque i. El subíndice h cuenta el número de empates en cada bloque. Por ejemplo, en el bloque 2 se tiene un solo empate (h = 1) y hay dos datos empatados (t_{2h} = 2). Luego $T_2 = 2^3 - 2 = 6$. Lo mismo ocurre para los bloques 5, 6 y 8, de modo que $\Sigma T_i = 24$, $C = 1 - 24/(24 \times 8) = 7/8$ y S corregido = $5.81 \times 8/7 = 6.64$: un valor que resulta mayor que $\chi^2_{0.05,2} = 5.99$.

En conclusión, las sustancias químicas A, B y C, tienen diferente grado de toxicidad (p = 0.036); la estadística S no corregida es no significativa: la corrección la convirtió en significativa, lo que muestra la importancia de la corrección cuando el valor de S está cerca del correspondiente valor crítico ($p \approx 0.040$ según tablas).

El análisis de comparaciones múltiples entre rangos para un diseño de bloques aleatorizados, sigue un procedimiento similar al de la sección 6.8. El valor crítico de referencia se calcula como el producto $W = q_{\alpha,k,\infty}\sqrt{kb(k+1)/12}$; toda diferencia entre dos sumas de rangos mayor que W se declara significativa con una rata de error experimental α. Esta prueba es aproximada pero mejora al aumentar el número de bloques.

Para $q_{0.05,3,\infty} = 3.314$, $W = 3.314\sqrt{3\times8\times4/12} = 9.4$. Entonces:

$R_2 - R_1 = 21.5 - 14.0 = 7.5 < 9.4$, diferencia no significativa.
$R_1 - R_3 = 14.0 - 12.5 = 1.5 < 9.4$, diferencia no significativa.
$R_2 - R_3 = 21.5 - 12.5 = 9.0 < 9.4$, diferencia no significativa.

Ninguna de las diferencias es significativa a pesar de que la prueba global corregida presentó una significancia leve. En primer lugar, tanto la prueba χ^2 para rangos como la de comparaciones múltiples son aproximadas, lo que explica el resultado anterior, el cual suele presentarse con alguna frecuencia en las investigaciones experimentales. Esta última prueba exige un número grande de bloques para mejorar la potencia; con ocho bloques la sensibilidad es baja y la conclusión final es que ocho ratas no dan la suficiente información para que este método detecte diferencias significativas. En segundo lugar, la prueba χ^2 es una aproximación de la prueba F cuando se usan rangos en vez de los datos originales, y esta aproximación solo es aceptable cuando $k > 7$, (Gibbons, 1993).

Tabla 7.14 Datos y sus rangos para el ejemplo 7.5

Ratas	*Sustancias tóxicas*		
	A	*B*	*C*
1	6 (3)	5 (2)	3 (1)
2	9 (2.5)	9 (2.5)	4 (1)
3	6 (2)	9 (3)	3 (1)
4	5 (1)	8 (3)	6 (2)
5	7 (1)	8 (2.5)	8 (2.5)
6	5 (1.5)	7 (3)	5 (1.5)
7	6 (2)	7 (3)	5 (1)
8	6 (1)	7 (2.5)	7 (2.5)
R_j	14.0	21.5	12.5

Ejercicios

7.1 Mediante un diseño experimental de bloques se comparó el número promedio de linfocitos (en miles/mm^3) en ratones, a los cuales se les administraron dos drogas experimentales y un placebo. Se seleccionaron ratones de la misma camada y del mismo sexo para formar los bloques (Gill, 1978).

Tratamientos	*Bloques*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Droga A	6.0	4.8	6.9	6.4	5.5	9.0	6.8
Droga B	5.1	3.9	6.5	5.6	3.9	7.0	5.4
Placebo	5.4	4.0	7.0	5.8	3.5	7.6	5.5

a. Establecer el modelo matemático apropiado para el análisis de estos datos.
b. Chequear los supuestos del modelo:
— Homogeneidad de varianzas para tratamientos y para bloques
— Análisis de residuos para normalidad y aditividad
Aplicar, si es necesario, la prueba de aditividad de Tukey.
c. Construir la tabla del ANOVA y obtener las conclusiones respectivas.
d. Comparar valores promedio de los tratamientos usando el procedimiento de Dunnett.
e. Discutir la efectividad de la formación de los bloques en este experimento.

7.2 Los datos de la siguiente tabla corresponden a un experimento realizado para comparar varios tratamientos relacionados con la fertilidad de semillas de fríjol soya. Los datos son el número de plantas que no emergieron de cada 100 que se plantaron por parcela (Snedecor y Cochran, 1984).

Bloques	*Control*	*Arasan*	*Spergon*	*Semesan, Jr.*	*Fermate*
1	8	2	4	3	9
2	10	6	10	5	7
3	12	7	9	9	5
4	13	11	8	10	5
5	11	5	10	6	3

a. Establecer el modelo matemático y las hipótesis del problema.
b. Construir la tabla del ANOVA. Chequear los supuestos.
c. Comparar el control con los demás tratamientos.
d. Comparar las medias de los tratamientos químicos, excluyendo el control.
e. Calcular la eficiencia del diseño.

7.3 El siguiente experimento se diseñó para comparar tres vitaminas. Para ello, se tomaron cuatro conjuntos de trillizos de un año de edad aproximadamente. Cada niño de cada familia recibió al azar uno de cada tres regímenes vitamínicos por dos años. En la tabla se reportan los pesos (en kg) ganados por los niños.

	Vitaminas		
Familia	*A*	*B*	*C*
1	4.95	4.05	4.50
2	3.60	4.08	4.10
3	4.90	4.50	4.92
4	3.64	3.65	4.50

a. Especificar la población, el factor, los niveles, la variable de bloqueo, la variable de respuesta, las unidades experimentales y las repeticiones.

b. Realizar el ANOVA y obtener conclusiones acerca de las vitaminas.

c. Estimar el número de bloques necesarios para garantizar una potencia de 0.80.

7.4 Ortiz (1991) comparó varias dietas alimenticias en larvas de *C. macropomum* (cachama negra). Los tratamientos fueron las combinaciones de tres alimentos y dos densidades: nauplios de Artemia (NAS), flan de huevo (FH), y NAS + FH, con densidades de 10 larvas por litro y 20 larvas por litro. La siguiente tabla muestra los pesos promedios diarios (en mg) para larvas provenientes de una sola pareja de reproductores, y con peso inicial de aproximadamente 0.58 mg. El experimento se realizó en la estación piscícola de Corpourabá en Choromandó, departamento de Antioquia (Colombia).

Tratamientos	*Días*							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
NAS10	0.84	0.90	1.10	1.20	1.40	2.20	2.60	2.80
FH10	0.65	0.80	0.80	0.60	0.61	0.80	0.65	0.65
NAS10 + FH10	0.80	1.00	1.10	1.40	1.00	1.20	1.70	1.50
NAS20	1.00	1.00	0.95	1.20	1.00	2.10	1.50	2.00
FH20	0.83	0.75	0.96	0.80	0.05	0.90	0.42	0.40
NAS20 + FH20	0.72	0.81	1.10	0.90	0.60	1.25	0.80	1.10

a. Tomar los días como bloques para hacer el análisis de estos datos. El valor 0.05 en el quinto día es un dato posiblemente mal registrado y deberá desecharse.

b. Estimar el dato que se desechó como un dato faltante.

c. Analizar la tabla del ANOVA correspondiente al diseño.

d. Comparar pesos promedio para densidades.

e. Comparar dietas para cada densidad,

f. Efectuar comparaciones por pares y discutir el resultado, comparándolo con los resultados en ***d*** y ***e***.

7.5 Para comparar tres metodologías en la determinación de los niveles de protrombina, se seleccionaron al azar ocho donantes que se consideraron como bloques. Los datos (en unidades / mL) aparecen en la tabla siguiente.

Método	*Donantes*							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
A	49	84	75	91	84	81	49	58
B	52	86	85	95	91	86	65	60
C	50	75	72	92	87	81	52	68

a. Establecer el modelo apropiado para analizar este diseño.

b. Efectuar el ANOVA y discutir los supuestos requeridos.

c. Realizar una comparación múltiple entre métodos.

7.6 En una investigación sobre contaminación ambiental, se seleccionó un área urbana y se midieron las concentraciones de óxido de nitrógeno y de hidrocarburos (en mg/m^3) tomando una muestra diaria durante varios días. Los datos fueron:

Días:	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
N_2O	104	116	84	77	66	81	72	68	97	84	90
Hidrocarburos	108	118	89	71	61	88	76	61	96	81	100

a. Considerar los días como bloques y establecer el modelo apropiado para el análisis.

b. Probar la hipótesis de que ambos contaminantes del aire están presentes en concentraciones similares.

c. Analizar los supuestos del modelo.

7.7 Se llevó a cabo un experimento para comparar el efecto de cuatro medicamentos sobre el tiempo (en minutos) de coagulación de la sangre. Se eligieron al azar diez personas sanas y se les administró, en orden aleatorio, los cuatro medicamentos a intervalos suficientes para eliminar efectos residuales. Cada persona se consideró como un bloque; de esta manera proporcionaron una lectura para cada medicamento. Los resultados se muestran en la tabla siguiente (adaptado de Daniel, 2004).

Medicamentos	*Personas*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
W	1.5	1.4	1.8	1.3	2.0	1.1	1.5	1.5	1.2	1.5
X	1.8	1.4	1.6	1.2	2.1	1.0	1.6	1.5	1.0	1.6
Y	1.7	1.3	1.5	1.2	2.2	1.0	1.5	1.5	1.3	1.6
Z	1.9	1.5	1.9	1.4	2.3	1.2	1.7	1.7	1.5	1.9

a. Establecer el modelo y sus supuestos para analizar estos datos.

b. Efectuar el ANOVA y obtener conclusiones.

c. Comparar los promedios de los medicamentos.

d. Estimar el número de personas necesarias en un experimento similar que garantice una potencia no inferior a 0.80.

7.8 En el ejercicio 5.7, supóngase que se eligieron al azar nueve individuos y que cada uno fue sometido a los tres métodos estudiados en fechas diferentes.

a. Formular el modelo matemático y efectuar el ANOVA correspondiente.

b. Comparar los resultados obtenidos con aquellos del ejercicio 5.7.

c. Calcular la eficiencia relativa del diseño de bloques con respecto al diseño inicial.

7.9 Se eligieron al azar diez animales de laboratorio para comparar las concentraciones arteriales (en ng/mL) de epinefrina en el plasma, durante tres tipos diferentes de anestesia. Los resultados fueron (Daniel, 2004):

Anestesia	*Animales*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
A	0.28	0.50	0.68	0.27	0.31	0.99	0.26	0.35	0.38	0.34
B	0.20	0.38	0.50	0.29	0.38	0.62	0.42	0.87	0.37	0.43
C	1.23	1.34	0.55	1.06	0.48	0.68	1.12	1.52	0.27	0.35

a. Especificar el modelo y su análisis correspondiente.

b. Efectuar el ANOVA y obtener conclusiones.

c. Comparar los efectos promedio de las anestesias.

7.10 En la siguiente tabla se muestran los porcentajes de mazorcas de maíz no comerciables que se obtuvieron en un experimento diseñado para comparar un testigo *A* y tres métodos mecánicos para protección contra daños de larvas del gusano de maíz. Los bloques están formados por parcelas a las que se les aplicó un solo tratamiento (Snedecor y Cochran, 1984).

Tratamientos	*Bloques*					
	1	*2*	*3*	*4*	*5*	*6*
A	42.4	34.3	24.1	39.5	55.5	49.1
B	33.3	33.3	5.0	26.3	30.2	28.6
C	8.5	21.9	6.2	16.0	13.5	15.4
D	16.6	19.3	16.6	2.1	11.1	11.1

a. Aplicar la prueba de Friedman para comparar los tratamientos.
b. Comparar tratamientos por pares usando rangos.

7.11 Livesey (1967), citado por Hollander y Wolfe (1999), comparó el desempeño de ratas, conejos y gatos en la prueba llamada de la vía elevada de Hebb-Williams. Los siguientes datos representan valores promedio de puntajes de error para doce problemas.

Especies	*Problemas*											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Ratas	1.5	1.1	1.8	1.9	4.3	2.0	8.4	6.6	2.4	6.5	2.6	6.5
Conejos	1.7	1.5	8.1	1.3	4.0	4.6	4.0	5.1	2.5	6.9	2.5	6.8
Gatos	0.3	1.0	3.6	0.0	0.6	5.5	1.0	3.1	0.1	1.6	4.3	1.0

a. Comparar especies usando la prueba de rangos de Friedman.
b. Usar un método de comparaciones múltiples para los rangos.
c. Describir las conclusiones de este experimento.

8

Diseños de cuadrado latino

8.1 Diseños de cuadrado latino

El diseño de bloques, definido en el capítulo anterior, se fundamenta en la estratificación de las unidades experimentales según una característica de variación conocida, con el fin de aislar dicha variación del error experimental. De este modo, un DBA resulta más eficiente que un DCA en la comparación de tratamientos o de niveles del factor en estudio.

El principio anterior puede extenderse un paso más mediante la eliminación de una segunda fuente de variación que conduzca a mejorar el DBA. Una de las fuentes de variación se elimina formando bloques horizontales y la otra bloques verticales.

La doble agrupación de las *ue* en bloques, y la correspondiente asignación al azar de los tratamientos, recibe el nombre de *diseño de cuadrado latino*, DCL. Se llama cuadrado latino porque al representar los k tratamientos por letras latinas, estas se distribuyen formando un cuadrado de $k \times k$ letras o de k^2 celdas, conocido como cuadrado latino de orden k. Para cuatro tratamientos A, B, C y D, el siguiente arreglo es un cuadrado latino de orden 4:

C	A	D	B
D	C	B	A
A	B	C	D
B	D	A	C

Para que k letras ordenadas en una matriz $k \times k$ sean un cuadrado latino, se requiere que cada letra o tratamiento esté una sola vez en cada fila y una sola vez en cada columna de la matriz. Lo anterior también se requiere para obtener estimaciones insesgadas de los efectos de los tratamientos. Esta restricción exige que las dos variables de control tengan igual número de niveles que el factor. Si el factor tiene k niveles, entonces se requieren k bloques horizontales y k bloques verticales para formar el cuadrado $k \times k$.

El DCL se usó originalmente en agricultura para eliminar de un conjunto de parcelas dispuestas en un cuadrado, los efectos de gradientes de fertilidad natural del suelo o los factores climáticos que podrían influir en dos direcciones diferentes a través del terreno cultivado.

La descripción del diseño como un cuadrado es conveniente, aunque las variables de control no sean factores físicos sino lógicos como la raza, la edad, el peso o el tiempo. En áreas como la medicina y la zootecnia, las personas y animales hacen las veces de bloques cuando reciben todos los tratamientos en secuencia; de igual manera, los periodos entre un tratamiento y otro forman los bloques de la otra variable, conformando un diseño de cuadrado latino. En general, se habla de filas y columnas para indicar los bloques horizontales y verticales respectivamente, y no se hacen diferencias entre los tipos de variables de control.

El DCL tiene la ventaja de que elimina dos fuentes de variación a la vez y también ahorra recursos, puesto que solo requiere k^2 *ue* en vez de k^3, como el diseño factorial de tres factores. Sin embargo, su utilidad está limitada a la presencia de factores que no interaccionen entre sí o con la variable de interés.

Por otro lado, aquellos factores cuyos efectos van a ser eliminados del error experimental deben estar relacionados con la variable bajo estudio, y su influencia sobre esta variable de respuesta debe ser reconocida. La eficiencia del diseño depende de la variación causada por los factores de control; si uno de ellos no ejerce una influencia apreciable en la variable de estudio, deberá adoptarse un DBA en vez de un DCL.

No todos los arreglos de letras en forma de cuadrado son cuadrados latinos. El número posible de cuadrados latinos $k \times k$ depende del valor de k, como se muestra a continuación:

Número de tratamientos	2	3	4	5	6
Total de cuadrados latinos	2	12	576	155 280 (miles)	812 851 200 (millones)

La asignación aleatoria de los k tratamientos a las k^2 *ue*, consiste en seleccionar al azar un cuadrado entre el total de cuadrados de orden k. En la práctica, esta selección aleatoria no es posible, pues no existen tablas que contengan todos los cuadrados latinos. Para aleatorizar un cuadrado latino, primero se obtiene un cuadrado estándar de una tabla de cuadrados —como las de Cochran y Cox (1995)— o se construye este, por ejemplo, colocando las letras en orden alfabético en cada fila; luego se aleatorizan filas, columnas y tratamientos por separado (Neter y otros, 1996).

8.2 Análisis de un diseño de cuadrado latino

En un diseño de cuadrado latino hay tres factores: dos que son variables de control y el factor de tratamientos. Los datos de un DCL se analizan mediante un ANOVA de tres vías con una observación por celda, suponiendo que todas las interacciones entre factores son insignificantes. El modelo matemático para un DCL con tratamientos y bloques fijos, es una extensión del modelo (7.1) para un DBA. Así:

$$y_{ijl} = \mu + \alpha_i + \tau_j + \beta_l + \varepsilon_{ijl}, \text{ con } i, j, l = 1, 2, \ldots, k, \qquad (8.1)$$

donde α_i es el efecto fijo de la fila i, con $\Sigma\alpha_i = 0$; β_l es el efecto fijo de la columna l, con $\Sigma\beta_l = 0$; τ_j es el efecto fijo del tratamiento j, con $\Sigma\tau_j = 0$ y ε_{ijl} es el efecto residual que incluye el error experimental y pequeños componentes de las interacciones. Se supone que los errores son independientes y normales $N(0,\sigma^2)$.

El modelo es aditivo. No hay interacciones entre filas y columnas y tampoco entre estas y los tratamientos; si existen son tan pequeñas que pueden igualarse a cero. De otro lado, el ANOVA consiste en subdividir la suma total de cuadrados en una componente para filas, otra para columnas, otra para tratamientos y otra para el error experimental, como se muestra en la tabla 8.1.

Para el modelo aleatorio cambian las medias cuadráticas esperadas, pero la prueba F es la misma que en la tabla anterior. Los *gl* del error son $(k - l)(k - 2)$ de modo que para $k = 2, 3, 4$ y 5, los *gl* respectivos son 0, 2, 6 y 12, por lo que un experimento con menos de cuatro tratamientos, con frecuencia no proporciona suficiente información para potencia y precisión adecuadas.

Ejemplo 8.1 Para comparar el efecto de cinco dietas, estas se administraron a cinco camadas, de cinco ratas cada una, mediante un diseño de cuadrado latino con camadas como filas y ratas ordenadas por peso como columnas. De esta manera se eliminaron las variaciones asignables a las diferencias en los pesos y en la herencia. La variable de respuesta fue el

Tabla 8.1 Tabla del ANOVA para un diseño de cuadrado latino

Fuente de variación	*SC*	*gl*	*MC*	*MC esperada*	*F*
Tratamientos	SCA	$k-1$	MCA	$\sigma^2 + k\Sigma\tau_j^2/(k-1)$	$\frac{MCA}{MCE}$
Filas	SCF	$k-1$	MCF	$\sigma^2 + k\Sigma\alpha_i^2/(k-1)$	
Columnas	SCC	$k-1$	MCC	$\sigma^2 + k\Sigma\beta_l^2/(k-1)$	
Error	SCE	$(k-1)(k-2)$	MCE	σ^2	
Total	SCT	k^2-1			

Tabla 8.2 Datos y cuadrado latino correspondientes al ejemplo 8.1

Camadas	*Peso en orden creciente*					*Total* $y_{i..}$
	1	*2*	*3*	*4*	*5*	
1	736(C)	587(E)	811(B)	724(A)	762(D)	3620
2	757(A)	737(C)	578(D)	590(E)	809(B)	3471
3	573(B)	789(D)	577(E)	751(C)	822(A)	3512
4	686(E)	761(B)	667(A)	683(D)	563(C)	3360
5	942(D)	802(A)	920(C)	635(B)	725(E)	4024
Total $y_{..l}$	3694	3676	3553	3383	3681	17987
Tratamientos	A	B	C	D	E	
Total $y_{.j.}$	3772	3589	3707	3754	3165	
Medias $\bar{y}_{.j.}$	754	718	741	751	633	

contenido de proteína en el hígado (mg/100 g de peso corporal). Los resultados se dan en la tabla 8.2 (adaptado de John y Quenouille, 1977).

Para construir la tabla del ANOVA se calculan las siguientes sumas:

$$C = \left(\frac{1}{k^2}\right)\left(\sum_i\sum_l y_{ijl}\right)^2 = \left(\frac{1}{25}\right)(17\,987)^2 = 12\,941\,287.$$

$$SCT = \sum\sum y_{ijl}^2 - C = (736)^2 + \cdots + (725)^2 - C = 267\,748.$$

$$SCC = \frac{1}{k}\sum_l y_{..l}^2 - C = \left(\frac{1}{5}\right)\left[(3694)^2 + \cdots + (3681)^2\right] - C = 14\,088.$$

$$SCF = \frac{1}{k}\sum_i y_{i..}^2 - C = \left(\frac{1}{5}\right)\left[(3620)^2 + \ldots + (4024)^2\right] - C = 52\,424.$$

$$SCA = \frac{1}{k}\sum_j y_{.j.}^2 - C = \left(\frac{1}{5}\right)\left[(3772)^2 + \ldots + (3165)^2\right] - C = 50\,812.$$

$$SCE = SCT - SCF - SCC - SCA = 150\,424.$$

Las camadas pueden considerarse como bloques aleatorios y los pesos como bloques fijos; esto no cambia la tabla del ANOVA. La prueba F nos dice que no hay diferencias significativas entre las dietas ($p = 0.44$). Igualmente, al observar las medias cuadráticas para filas y para columnas, se deduce que no hay significancia estadística. La conclusión es que ni las camadas ni el peso son factores influyentes en la variable de respuesta; por tanto, el diseño de cuadrado latino no es el mejor para este experimento.

No deben olvidarse los supuestos del modelo; para este ejemplo todos ellos se cumplen. En particular, la aditividad es aceptable de acuerdo con la gráfica de los residuos e_{ijl} contra $\hat{y}_{ijl}$. También existe la prueba de aditividad de Tukey para cuadrados latinos, esta se encuentra ejemplificada en Snedecor y Cochran (1984).

La eficiencia relativa del DCL con respecto al DBA, suponiendo que las camadas son los bloques, se calcula con la siguiente fórmula:

$$E_R = \frac{MCC + (k-1)MCE}{(k)MCE} = \frac{3522 + 4(12535)}{5(12535)} = 0.86.$$

Como se observa, no hay ganancia en la eficiencia por el hecho de considerar el peso de las ratas como otra variable de control. La eficiencia relativa del cuadrado latino con respecto a un DCA, indica que ninguna de las dos variables de bloqueo influye en forma apreciable sobre la respuesta analizada, por tanto, pueden no tenerse en cuenta en estudios futuros. Esta última eficiencia es:

$$E_R = \frac{MCC + MCF + (k-1)MCE}{(k+1)MCE} = \frac{3522 + 13106 + 4(12535)}{6(12535)} = 0.89.$$

Tabla 8.3 Tabla del ANOVA para los datos del ejemplo 8.1

Fuentes de variación	*SC*	*gl*	*MC*	*F*	*p*
Dietas	50 812	4	12 703	1.01	0.44
Filas	52 424	4	13 106		
Columnas	14 088	4	3 522		
Error	150 424	12	12 535		
Total	267 748	24			

Por causas diversas, como se mencionó en el DBA, pueden perderse uno o más datos del cuadrado latino. Este hecho destruye la ortogonalidad entre tratamientos y bloques, y las fórmulas del ANOVA dejan de ser aplicables. Si se tiene un solo dato faltante, se puede hacer una estimación de su valor mediante la fórmula:

$$\hat{y}_{ijl} = \frac{k\left(y'_{i..} + y'_{.j.} + y'_{..l}\right) - 2y'_{...}}{(k-1)(k-2)}.$$

y continuar con el análisis en forma análoga al ejemplo 7.3. El sesgo en la *SCA* se calcula como:

$$B = \left[\frac{y'_{...} - y'_{i..} - y'_{..l} - (k-1)\,y'_{.j.}}{(k-1)(k-2)}\right]^2.$$

El error estándar de la diferencia entre la media del tratamiento con el dato faltante y cualquiera otra media es:

$$EE = \sqrt{MCE\left[\frac{2}{k} + \frac{1}{(k-1)(k-2)}\right]}.$$

Dos o más datos faltantes hacen que el diseño sea menos ortogonal y, por tanto, las estimaciones son más complejas. El método de regresión puede aplicarse siempre, no importa cuántos sean los datos perdidos.

Teóricamente, podrían diseñarse experimentos controlando simultáneamente tres, cuatro o más variables. Para tres factores de control, el diseño se llama cuadrado greco-latino; para cuatro, cuadrado hipergreco-latino. Son diseños difíciles de construir y algunos ni siquiera existen, por ejemplo el greco-latino de orden 6 × 6. En las ciencias biológicas son poco conocidos; es en las ciencias sociales y en ingeniería donde tienen mayor aplicación.

8.3 Diseños de bloques incompletos

A veces el número de tratamientos es mayor que el número de *ue* homogéneas disponibles para formar bloques completos. Por ejemplo, el tamaño de la camada puede ser menor que el número de tratamientos; en un solo día no pueden aplicarse todos los tratamientos, suponiendo que cada día es un bloque, o, como ocurre en ensayos clínicos, las cohortes de pacientes —grupos de la misma edad, peso o estrato social— disponibles en un tiempo corto son generalmente insuficientes.

Si en un bloque el número de *ue* es menor que el de tratamientos, se dice que el bloque es incompleto; un diseño formado por tales bloques se llama *diseño de bloques incompletos*.

Ahora bien, sean k el número de tratamientos en el experimento y $t(t < k)$ el número de *ue* disponible por bloque. Para asegurar igual precisión en las comparaciones entre los diferentes pares de tratamientos, estos deben distribuirse en los bloques de tal forma que cada par y cada tratamiento tengan igual número de repeticiones. Los diseños que satisfacen esta condición reciben el nombre de *diseños de bloques incompletos balanceados*.

La construcción de estos diseños no siempre es fácil, pues se requiere el conocimiento de teorías matemáticas complejas como la de las geometrías finitas. F. Yates fue el pionero en su desarrollo y los aplicó para experimentos agrícolas alrededor de 1930. Posteriormente, R.C. Bose, K. R. Nair y otros miembros de la escuela estadística Hindú, construyeron diseños más avanzados, conocidos como *diseños de bloques incompletos parcialmente balanceados*.

Como ejemplo, un diseño para comparar cinco tratamientos con solo tres tratamientos por bloque y que sea balanceado requiere diez bloques: ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE.

En este diseño, cada tratamiento está repetido seis veces y cada par de tratamientos tres veces; es decir, dos tratamientos aparecen juntos en tres bloques. Ejemplos que ilustran la construcción y análisis de diseños de bloques incompletos se encuentran en Cochran y Cox (1991), Cox (1992), Hicks y Turner (1999) y Montgomery (2007).

En el caso de diseño de bloques incompletos balanceados para los que el número de tratamientos es igual al número de bloques ($k = b$), es posible arreglar los tratamientos dentro de los bloques de manera que formen un cuadrado latino incompleto. Estos diseños se conocen como cuadrados de Youden —en realidad son rectángulos— porque fue W. J. Youden el primero en publicarlo en 1940.

Si en un cuadrado latino falta una fila o una columna, pero no ambas, el diseño pierde la ortogonalidad: es un cuadrado latino incompleto, pero puede analizarse mediante un ANOVA de bloques incompletos, o puede ignorarse una de las variables de bloqueo y analizarse como un DBA. Si se desea un análisis más riguroso, deberá emplearse el método de regresión. Aun faltando una fila y una columna, o un bloque y un tratamiento, el diseño sigue teniendo validez y es posible analizarlo estadísticamente.

8.4 Diseños conmutativos

El diseño de medidas repetidas más simple es el que asigna los tratamientos al azar y en secuencia al mismo sujeto. Este diseño puede usarse únicamente cuando los sujetos responden en forma más o menos similar a los tratamientos a través del tiempo, esto es, no hay tendencias asignables al sujeto, y también cuando los tratamientos se dan suficientemente espaciados de modo que no haya efectos residuales entre dos tratamientos consecutivos.

Si en el experimento se sospecha algún tipo de tendencia en los sujetos, es preferible asignar los tratamientos en un orden balanceado en vez de aleatorio. Para tres tratamientos A, B y C, una secuencia balanceada sería: ABC, BCA, CAB. La aleatorización del diseño consiste en distribuir los b sujetos al azar y en igual número, a las secuencias escogidas para el experimento. El orden balanceado no debe usarse cuando los tratamientos tienen efectos residuales, ya que en este diseño un tratamiento particular siempre está seguido del mismo tratamiento en la secuencia.

Los diseños conmutativos, llamados también diseños con intercambio (*cross-over* o *change-over*), buscan eliminar la heterogeneidad entre individuos. Para ello se toman

los sujetos como bloques, siempre que los tratamientos sean de tal naturaleza que pueden aplicarse todos al mismo sujeto, desde luego, en tiempos diferentes.

El modelo matemático para un diseño de bloques conmutativo no es el mismo modelo (7.2), pues es necesario tener en cuenta el efecto de la secuencia de los tratamientos o efecto de periodo. El modelo apropiado es de la forma:

$$y_{ijl} = \mu + B_i + \tau_j + \delta_l + \varepsilon_{ijl};\ \text{con } i = 1, 2, \ldots, r\ j, l = 1, 2, \ldots, k. \quad (8.2)$$

Las interacciones entre los efectos de sujetos B_i, tratamientos τ_j y de periodos δ_l, se consideran nulos o insignificantes. El error ε_{ijl} refleja la variación aleatoria de la información repetida dentro de los sujetos, como si los tratamientos no fueran impuestos y no existiera efecto de periodo.

Los cálculos para el ANOVA son similares a los del cuadrado latino, pero con una estructura como la de la tabla 8.4. Además el número de secuencias diferentes es igual al número de tratamientos, pero el número de bloques puede ser un múltiplo del de tratamientos.

Tabla 8.4 Tabla del ANOVA para un diseño de bloques conmutativo con efectos de periodos

Fuente de variación	*SC*	*gl*	*MC*	*F*
Tratamientos	SCA	$k-1$	MCA	$\frac{MCA}{MCE}$
Sujetos	SCS	$r-1$	MCS	
Periodos	SCP	$k-1$	MCP	
Error	SCE	$(k-1)(r-2)$	MCE	
Total	SCT	$kr-1$		

Ejemplo 8.2 Para comparar el efecto de dos formulaciones de una droga, se seleccionaron al azar dieciocho personas de una población. Las dieciocho personas se dividieron, nuevamente al azar, en dos grupos de nueve: al primer grupo se le dio la formulación A y al segundo la formulación B. Después de ocho horas, a cada persona se le registró la cantidad de droga acumulada en la sangre. Transcurrido un tiempo prudencial, para evitar posibles efectos residuales de la droga, se repitió el experimento, dándole al primer grupo la formulación B y al segundo la A. Los datos obtenidos aparecen en la tabla 8.5 (adaptado de Anderson y McLean, 1974).

Con las siguientes sumas de cuadrados se construyó la tabla 8.6.

$C = (1/36)(1673.4)^2 = 77785.2$

$SCT = (51.9)^2 + (35.1)^2 + \ldots + (58.1)^2 - C = 3554.7$

$SCA = [(1/18)[(830.3)^2 + (843.1)^2] - C = 4.6$

$SCP = [(1/18)[(683.7)^2 + (989.7)^2] - C = 2601.0$

$SCS = [(1/2)[(110.4)^2 + (95.5)^2 + \ldots (92.4)^2] - C = 619.2$

$SCE = 3554.7 - 4.6 - 2601.0 - 619.2 = 329.9.$

No hay diferencias significativas entre las dos formulaciones de la droga ($p = 0.65$). También puede observarse la gran variación entre periodos en el experimento, lo que indica el éxito en el diseño conmutativo. Además, con el fin de ampliar el campo inferencial, los sujetos fueron seleccionados al azar de una población mayor, sin embargo, los resultados muestran que la variación entre ellos no es significativa ($p = 0.13$).

Las medias de los tratamientos y sus errores estándares se calculan mediante:

$\bar{y}_{.j.} \pm \sqrt{MCE/r}$.

Para la formulación A el resultado es $830.3/18 \pm \sqrt{31.6/18} = 46.1 \pm 1.3$, y para la formulación B es 46.8 ± 1.3; por tanto, no hay diferencias entre las dos formulaciones de la droga.

Por otra parte, la estimación de los datos faltantes en este diseño sigue el procedimiento que se empleó en el diseño de cuadrado latino, como lo explica Li (1969a).

8.5 Cuadrados latinos repetidos

Los cuadrados latinos grandes, por ejemplo aquellos con nueve o más tratamientos, son poco utilizados porque requieren demasiadas *ue* que normalmente no pueden obtenerse. De otro lado, si se tiene pocos tratamientos, por ejemplo tres o cuatro, los *gl* del error no son suficientes para garantizar una precisión aceptable. En este caso, para que el diseño sea efectivo, puede planearse un experimento para ejecutar varios cuadrados latinos del mismo orden. Los tratamientos serán los mismos en cada

Tabla 8.5 Datos para el ejemplo 8.2. Tratamientos y periodos fijos; bloques, en este caso sujetos, aleatorios

Sujeto	*1(A)*	*2(B)*	*Total*	*Sujeto*	*1(B)*	*2(A)*	*Total*
1	51.9	58.5	110.4	10	50.8	62.2	113.0
2	35.1	60.4	95.5	11	41.1	51.4	92.5
3	38.6	50.4	89.0	12	39.1	50.7	89.9
4	36.1	58.7	94.8	13	35.7	51.1	88.6
5	34.6	64.8	99.4	14	33.7	51.4	85.1
6	39.7	54.9	94.6	15	31.2	45.1	76.3
7	37.8	56.4	94.2	16	34.3	51.1	85.4
8	38.8	52.5	91.3	17	31.8	57.5	89.3
9	39.1	54.5	93.6	18	34.3	58.1	92.4
Sumas	351.7	511.1			332.0	478.6	

Total periodo 1: 351.7 + 332.0 = 683.7
Total periodo 2: 511.1 + 478.6 = 989.7
Total formulación A: 351.7 + 478.6 = 830.3
Total formulación B: 511.1 + 332.0 = 843.1

Tabla 8.6 Tabla del ANOVA para el ejemplo 8.2

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Formulaciones	4.6	1	4.6	0.22	0.65
Sujetos	619.2	17	36.4	1.77	0.13
Periodos	2601.0	1	2601.0	126.1	<0.001
Error	329.9	16	20.6		
Total	3554.7	35			

	Sujeto 1	Sujeto 2
Periodo I	A	B
Periodo II	B	A

Figura 8.1 Cuadrado latino para dos secuencias y dos sujetos

cuadrado, pero la aleatorización será diferente, y además podrían usarse o no las mismas *ue*, las mismas filas y las mismas columnas. En cada caso, la subdivisión de los *gl* es diferente. Box, Hunter y Hunter (1993). Cada cuadrado latino se analiza separadamente y al final se suman las correspondientes sumas de cuadrados, obtenidas de los diferentes cuadrados, para construir la tabla del ANOVA.

Otra aplicación de los cuadrados latinos repetidos se encuentra en los diseños conmutativos. En estos diseños, a cada sujeto se le asigna una secuencia que, para el ejemplo 8.2, puede ser AB o BA. La aplicación de secuencias diferentes a individuos diferentes origina una serie de cuadrados latinos requiere un análisis apropiado de los datos. Para dos secuencias y dos sujetos, se tiene un cuadrado como el de la figura 8.1

Con dieciocho sujetos, por ejemplo, pueden construirse nueve cuadrados latinos 2 × 2. Aquí, el diseño conmutativo combina las características del diseño de bloques pequeños con los cuadrados latinos pequeños. Hay muchas formaciones posibles de bloques en los diseños conmutativos; el enfoque apropiado depende del número de tratamientos y del número de periodos en cada secuencia de tratamientos. Amplias ilustraciones de estos diseños se encuentran en Li (1969a) y en Cochran y Cox (1995).

El estudio de efectos residuales en diseños con medidas repetidas tomando los sujetos como bloques, es otro ejemplo de aplicación de los cuadrados latinos completos e incompletos. Se puede observar que, mediante la combinación de algunos diseños básicos, se construyen diseños más complejos que permiten resolver problemas de creciente dificultad, por ejemplo en estudios de biodisponibilidad para establecer bioequivalencia entre productos farmacéuticos.

Ejercicios

8.1 Con el fin de comparar varios sistemas de pesca, se diseñó un experimento en cuadrado latino, tomando como variables de control las horas del día y la habilidad de los pescadores. Los tratamientos A, B, C y D fueron cuatro clases de palangres usados por cuatro pescadores —cada pescador utilizaba un palangre durante tres horas, luego se rotaban éstos hasta que todos hubiesen usado cada palangre— en una jornada de trabajo de 12 horas. La variable de respuesta fue el número de peces capturados por pescador usando un solo palangre (Bazigos, 1976).

Pescadores	*Periodo de pesca*			
	1	*2*	*3*	*4*
1	A:15	B:10	D: 8	C: 8
2	C:16	D:17	B:10	A:11
3	B:13	C:11	A:12	D: 7
4	D:15	A:10	C:10	B: 9

a. Establecer el modelo para analizar estos datos y sus supuestos,
b. Enunciar las hipótesis estadísticas del problema.
c. Efectuar el ANOVA respectivo y obtener las conclusiones del caso.
d. Calcular la eficiencia relativa del diseño aplicado según las conclusiones en *c*.
e. Describir otra aleatorización posible para este mismo diseño.

8.2 Se utilizó un diseño de cuadrado latino para comparar cuatro dosis de insulina A, B, C y D, aplicadas a conejos, a los que se les midió el contenido de azúcar en la sangre (en mg de glucosa/100 cm^3 de sangre) pocas horas después de haber sido inyectados; las variables de bloques fueron los conejos y los días de aplicación. El efecto de la inyección de insulina desaparece en pocos días sin dejar efectos residuales en el animal. En la siguiente tabla se dan los datos obtenidos en este estudio (Finney, 1963).

Días	*Conejos*			
	1	*2*	*3*	*4*
1	B:47	A:90	C:79	D:50
2	D:46	C:74	B:63	A:69
3	A:62	B:61	D:58	C.66
4	C:76	D:63	A:87	B:59

a. Formular un modelo adecuado para analizar estos datos. Tener en cuenta que el efecto columna es aleatorio.
b. Construir la tabla del ANOVA y obtener las conclusiones respectivas.
c. Comparar las cuatro dosis de insulina. Se sabe que A y C son dosis más bajas que B y D.
d. Estimar la variabilidad entre animales experimentales, considerados como una muestra de una población de conejos de laboratorio.
e. Calcular la eficiencia relativa de este diseño.

8.3 Se diseñó un experimento conmutativo para estudiar la respuesta de la prolactina en el suero después de diferentes lavados de la ubre de vacas lecheras. Para ello se seleccionaron diez vacas en etapas similares de lactancia: cinco de ellas se asignaron a la secuencia AB y las restantes a la BA. Cada lavado duraba 30 segundos en intervalos de A: 5 minutos y B: 15 minutos, durante una hora. Los datos representan las diferencias en la prolactina medida cinco minutos después del último lavado y la prolactina medida antes de cada tratamiento (Gill, 1978, vol. 2).

Vacas	*1(A)*	*2(B)*	*Vacas*	*1(B)*	*2(A)*
1	19.5	10.7	6	8.9	11.9
2	8.3	2.2	7	61.4	68.2
3	78.2	70.5	8	40.2	46.0
4	10.5	2.2	9	13.4	17.8
5	38.2	33.8	10	4.2	9.4

a. Construir la tabla del ANOVA para estos datos.

b. Comparar la efectividad promedio de los dos procedimientos de lavado.
c. Discutir el efecto de periodos.
d. Discutir el efecto de bloques, es decir, la variabilidad entre vacas.

9

Experimentos factoriales

9.1 Caracterización de los experimentos

De acuerdo con Das y Giri (1986), hay tres tipos fundamentales de experimentos que requieren investigación estadística: experimentos con tratamientos, experimentos factoriales y bioensayos.

En los primeros, los tratamientos son generalmente los diferentes niveles de un factor, por ejemplo, diferentes variedades de semilla, tipos de alimentos para animales, dosis crecientes de una droga, etc. El objetivo de estos experimentos es comparar los tratamientos por grupos. Este es el tipo de experimentos que se ha analizado en los capítulos anteriores.

En los experimentos factoriales, los tratamientos son combinaciones de dos o más niveles de factores diferentes. Por tanto, un experimento factorial involucra más de un factor, cada uno con dos o más niveles. En estos experimentos no se establecen comparaciones por pares sino un tipo diferente de comparaciones, llamadas efectos principales e interacciones.

Los experimentos llamados bioensayos —ensayos o experimentos biológicos— son experimentos para estimar la naturaleza, constitución o potencia de una sustancia, por medio de la reacción obtenida al aplicarla a material biológico. Lo usual es tomar dos preparaciones de drogas que tengan un efecto común; una de ellas se considera estándar y la eficacia de la otra se estima mediante el cociente entre diferentes dosis de las drogas. Los bioensayos requieren diseños muy específicos que no se estudian en un texto introductorio al diseño experimental (véase Govindarajulu, 2001).

En los experimentos factoriales, objeto de este capítulo, los niveles de los factores se estudian combinados y se llaman, usualmente, combinaciones de tratamientos, por ejemplo, un experimento para comparar cuatro variedades de semillas usando dos fertilizantes es un experimento factorial 2×4; lo que quiere decir que se toman dos niveles de un factor combinados con cuatro niveles del otro factor, para formar ocho combinaciones de tratamientos. Denotando los fertilizantes por f_0 y f_1 y las variedades por v_0, v_1, v_2, y v_3, se pueden formar las combinaciones siguientes: f_0v_0, f_0v_1, f_0v_2, f_0v_3, f_1v_0, f_1v_1, f_1v_2 y f_1v_3. En adelante, cada combinación se considerará simplemente como

un tratamiento. Si se analizan tres factores, dos a tres niveles y uno a cinco niveles, el experimento factorial será de la forma $3 \times 3 \times 5$ o $3^2 \times 5$.

Un tipo especial de experimento es el factorial simétrico que se caracteriza porque los factores tiene igual número de niveles. Los más utilizados son los experimentos 2^n y 3^n, donde n representa el número de factores. Si los factores tienen diferente número de niveles, el experimento es un factorial asimétrico.

Teóricamente es posible analizar experimentos con varios factores —cada uno con varios niveles—, pero razones prácticas hacen que el tamaño del experimento sea limitado. Por ejemplo, un experimento con tres factores, cada uno con cinco niveles, necesitaría $5^3 = 125$ *ue* u observaciones para una sola repetición del experimento. Existen técnicas especiales —confusión, diseños fraccionados, entre otras— para reducir el tamaño de un experimento factorial, de modo que con pocas *ue* puedan estudiarse varios factores a la vez. Por otra parte, un experimento factorial puede diseñarse completamente al azar, en bloques o en cuadrados latinos, y tiene las mismas ventajas y desventajas de estos diseños.

Los experimentos factoriales son más eficientes que los unifactoriales, porque permiten estudiar no solo efectos individuales sino también el efecto conjunto o combinado de los factores; a este efecto se le ha llamado *interacción*. En muchos problemas de investigación, la información obtenida acerca de los efectos combinados de los factores sobre la variable de respuesta, contribuye en mayor proporción a la solución del problema que la información obtenida de los efectos individuales o efectos principales de los mismos factores.

Box, Hunter y Hunter (2005) trae el siguiente ejemplo que ilustra con mucha claridad el concepto de interacción entre dos variables independientes:

> Consideremos un estudio experimental de los efectos del alcohol y del café en los tiempos de respuesta de los conductores de automóviles sentados en un simulador. Supongamos que se ha encontrado que: a) si no ha tomado café, un trago de licor aumenta el tiempo de respuesta en una media de 0.45 segundos y b) si no ha tomado alcohol, una taza de café reduce el tiempo de respuesta en una media de 0.20 segundos.
>
> Al evaluar los efectos de varias copas de licor y varias tazas de café y su efecto combinado, se podría simplificar mucho el trabajo si ambos efectos fueran lineales y aditivos. Si fueran lineales, dos copas de licor incrementarían el tiempo de respuesta en 0.90 segundos ($2 \times 0.45 = 0.90$) y tres tazas de café lo reducirían en 0.60 segundos ($-0.20 \times 3 = -0.60$). Si los efectos fueran aditivos, una copa de licor y una taza de café incrementarían el tiempo de reacción en 0.25 segundos ($0.45 - 0.20 = 0.25$). Finalmente, si fueran lineales y aditivos, 10 copas de licor y 23 tazas de café reducirían el tiempo de reacción en 0.10 segundos $(10 \times 0.45) - (23 \times 0.20) = -0.10$.
>
> Es mucho más probable, sin embargo, que el efecto de una copa más de licor dependa de a) el número de copas de licor que ya se haya bebido (el efecto del alcohol no es lineal) y b) el número de tazas de café que se haya tomado antes (existe un efecto de interacción entre el alcohol y el café).

9.2 Modelo matemático en estudios bifactoriales

En un experimento con dos factores debe tenerse en cuenta si ambos factores son de efectos fijos, de efectos aleatorios, o uno es de efectos fijos y el otro de efectos aleatorios. Si el diseño es completamente aleatorizado, el número de repeticiones por celda puede ser $r = 1$ o $r > 1$.

Considérese el caso del modelo I —efectos fijos— y un DCA con $r > 1$ repeticiones. Un factor A con a niveles y un factor B con b niveles conforman los ab tratamientos incluidos en el estudio, y el número de observaciones será abr. La observación k-ésima para el tratamiento donde el factor A está en el nivel i; y el B en el nivel i, es y_{ijk}. Un modelo matemático que incorpora la interacción entre factores es de la forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}; \text{ con i} = 1, 2, \dots, a,$$

$$j = 1, 2, \dots, b \text{ y } k = 1, 2, \dots, r. \quad (9.1)$$

donde μ es la media poblacional global, α_i es el efecto principal del nivel i del factor A, β_j es el efecto principal del nivel j del factor B, y $(\alpha\beta)_{ij}$ es el efecto de la interacción del nivel i de A con el nivel j de B.

Para estimar parámetros debe suponerse que

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0.$$

Además, que los errores ε_{ijk} son independientes, con distribución $N(0, \sigma^2)$ y que las varianzas σ^2_{ij} de las celdas son homogéneas. Los electos de interacción $(\alpha\beta)_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$ serán cero cuando los factores actúen aditivamente; pero algunos serán diferentes de cero cuando haya interacción entre los factores.

Normalmente, el interés del investigador está centrado en el efecto de la interacción, por eso el análisis estadístico inicia con la prueba F de no interacción. Las hipótesis son:

H_0: $(\alpha\beta)_{ij} = 0$ en todas las celdas.
H_1: $(\alpha\beta)_{ij} \neq 0$ en algunas de las celdas.

Si se concluye H_1, el paso siguiente es analizar las medias μ_{ij} de las celdas. Generalmente, no se prueba la significancia de los efectos principales de cada factor ya que habiendo interacción, estos efectos carecen de significado o su interpretación puede ser bastante compleja. En este caso, una de las mejores ayudas en la interpretación de los resultados es la gráfica de medias o totales de celda.

Si los efectos principales son aditivos, se prueba la significancia para cada factor mediante las hipótesis:

H_0: $\alpha_i = 0$ contra H_1: $\alpha_j \neq 0$ Para el factor A y
H_0: $\beta_j = 0$ contra H_1: $\beta_j \neq 0$ Para el factor B.

Los cocientes F para cada una de las pruebas se dan en la tabla 9.1, en la cual se resume el ANOVA para el diseño que se ha estado analizando.

Ejemplo 9.1 Dos razas de *Drosophila pseudoobscura* fueron producidas por endogamia para resistir un insecticida. Se analizaron cuatro niveles de concentración de insecticida en ambas razas. Los datos de la tabla 9.2, expresados en porcentajes de mortalidad durante un periodo determinado, están basados en tres repeticiones de cada tratamiento (adaptado de Schefler, 1981).

Tabla 9.1 Tabla del ANOVA para un diseño bifactorial, modelo I

Fuente de variación	*SC*	*gl*	*MC*	*MC esperado*	*F*
Factor A	SCA	$a-1$	MCA	$\sigma^2 + \frac{br}{a-1}\sum \alpha_i^2$	$\frac{MCA}{MCE}$
Factor B	SCB	$b-1$	MCB	$\sigma^2 + \frac{ar}{b-1}\sum \beta_j^2$	$\frac{MCB}{MCE}$
Interacción AB	SCAB	$(a-1)(b-1)$	MCAB	$\sigma^2 + \frac{r}{(a-1)(b-1)}\sum\sum(\alpha\beta)_{ij}^2$	$\frac{MCAB}{MCE}$
Error	SCE	$ab(r-1)$	MCE	σ	
Total	SCT	$abr-1$			

Para construir la tabla del ANOVA (véase tabla 9.3) se supondrá que el factor A es la raza y que el factor B es la concentración del insecticida, y se calculan las sumas de cuadrados como sigue:

$$C = \frac{1}{abr}\left[\sum_i\sum_j\sum_k y_{ijk}\right]^2 = \left(\frac{1}{24}\right)(1\,170)^2 = 57037.5.$$

$$SCT = \sum_i\sum_j\sum_k y_{ijk}^2 - C = 59456 - 57037.5 = 2418.5.$$

$$SCA = \frac{1}{br}\sum_i\left[\sum_j\sum_k y_{ijk}\right]^2 - C = \frac{1}{12}\left[(616)^2 + (554)^2\right] - C = 160.2.$$

$$SCB = \frac{1}{ar}\sum_j\left[\sum_i\sum_k y_{ijk}\right]^2 - C = \frac{1}{6}\left[(297)^2 + \ldots + (296)^2\right] - C = 179.5.$$

$$SCAB = \frac{1}{r}\sum_i\sum_j\left[\sum_k y_{ijk}\right]^2 - C - SCA - SCB.$$

$$= \frac{1}{3}\left[(167)^2 + \ldots + (142)^2\right] - C - SCA - SCB = 858.2.$$

$$SCE = SCT - SCA - SCB - SCAB = 1220.6.$$

La interacción entre los factores es significativa, lo que indica que el porcentaje de mortalidad varía según la raza y la concentración. Como se puede observar en la figura 9.1, construida con base en las medias de las celdas de la tabla 9.4, al pasar de la concentración C_1 a la concentración C_2, la mortalidad para la raza R_2 aumenta mientras que para la raza R_1 disminuye. Al pasar de C_2 a C_3, esta tendencia se invierte y, finalmente, al pasar de C_3 a C_4, la variación se asemeja más al primer paso que al segundo. En ausencia de interacción, las líneas mostrarían alto grado de paralelismo.

Con respecto a los supuestos del modelo se ha chequeado gráficamente la normalidad de los residuos y se ha encontrado que es aceptable. Es de anotar que los datos son porcentajes pero no requieren transformación, ya que su rango está entre el 30 y el 80%. Sin embargo, una transformación adecuada podría convertir el modelo con interacción en un modelo aditivo.

La homogeneidad de varianzas entre celdas se comprobó usando el método de Burr-Foster ($q = 0.15, p > 0.01$). Si la hipótesis de homogeneidad fuera rechazada, podría buscarse una transformación que estabilice las varianzas o también podrían considerarse

Tabla 9.2 Porcentajes de mortalidad según el ejemplo 9.1

Razas	*Concentraciones*				*Totales*
	C_1	C_2	C_3	C_4	
R_1	60	44	46	31	616
	55 (167)	37 (135)	51 (160)	57 (154)	
	52	54	63	66	
R_2	37	63	30	51	554
	43 (130)	59 (176)	38 (106)	50 (142)	
	50	54	38	41	
Totales	297	311	266	296	1170

Tabla 9.3 Tabla del ANOVA para el ejemplo 9.1

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Raza (A)	160.2	1	160.2	2.10	0.17
Insecticida (B)	179.5	3	59.8	0.78	0.52
Interacción (AB)	858.2	3	286.1	3.75	0.033
Error	1220.6	16	76.3		
Total	2418.5	23			

Tabla 9.4 Medias y errores estándar para los datos del ejemplo 9.1

Razas	*Concentraciones*				*Medias*
	C_1	C_2	C_3	C_4	
R_1	55.7	45.0	53.3	51.3	51.3
R_2	43.3	58.7	35.3	47.3	46.2
Medias	49.5	51.9	44.3	49.3	48.8

Errores estándar

1. De medias de celdas: $\sqrt{MCE/r} = \sqrt{76.3/3} = 5.04$

2. De medias de razas: $\sqrt{MCE/br} = \sqrt{76.3/12} = 2.52$

3. De medias de concentraciones: $\sqrt{MCE/ar} = \sqrt{76.3/6} = 3.57$

pruebas aproximadas para los efectos de los tratamientos. Las fórmulas para los *gl* de las pruebas *F* aproximadas, se encuentran en Gill (1978, vol. 1).

La tabla 9.3 muestra que hay algo de interacción, pero en realidad no es muy significativa ($p = 0.033$). En los datos originales no hay interacción; lo cual es de esperarse, ya que las concentraciones son variantes de una misma sustancia y las razas son similares entre sí. Los datos en estudio fueron escogidos intencionalmente para obtener interacción entre los factores, pues el ejemplo tiene como objetivo mostrar el análisis requerido cuando se presenta esta situación.

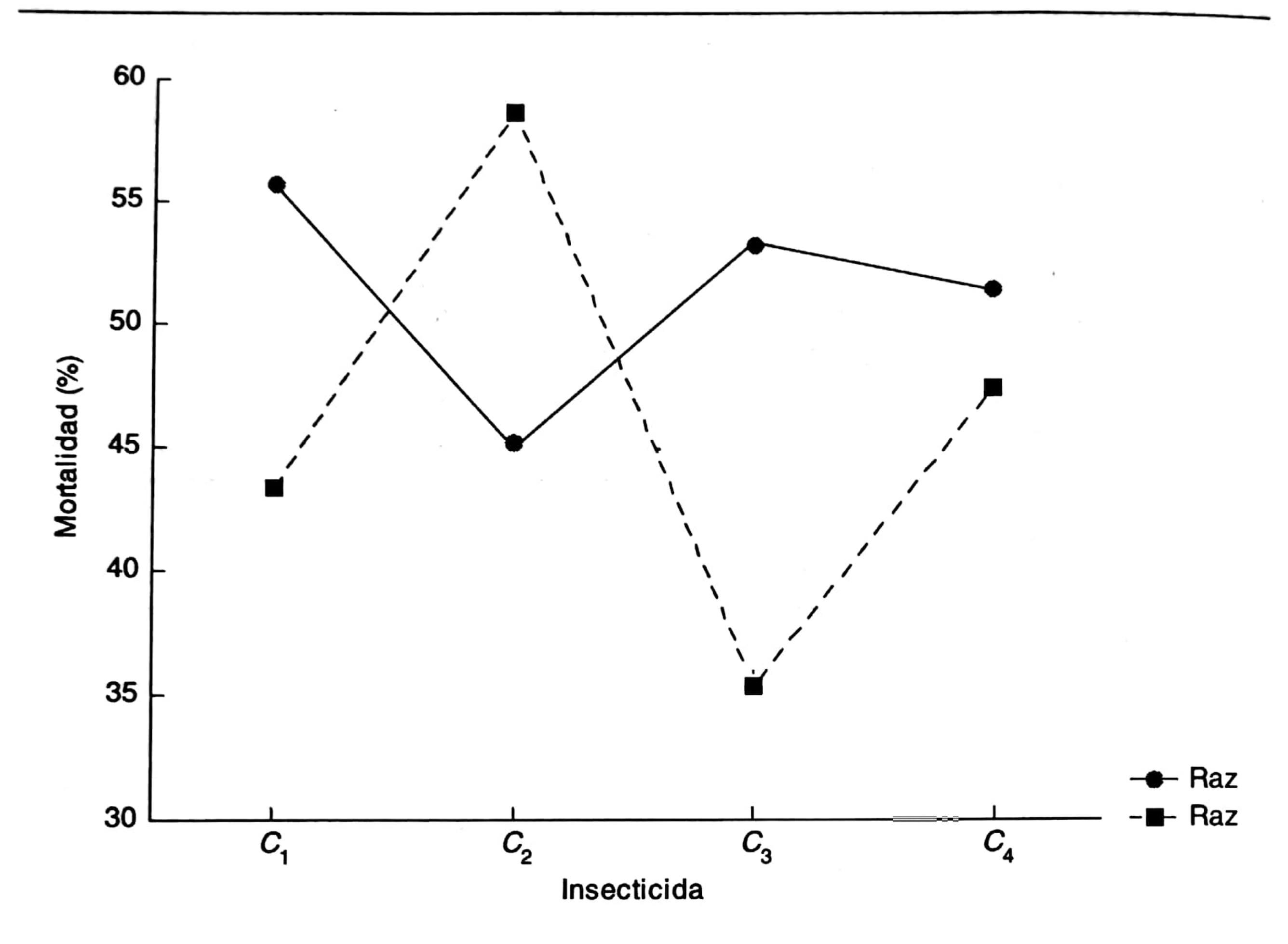


Figura 9.1 Interacción entre raza e insecticida

Las comparaciones entre medias dependen del interés del investigador. En el caso de interacción entre factores cualitativos, es usual investigar la combinación de niveles que produce la mejor respuesta. Para ello se supone que el modelo es de una vía con ocho niveles y tres repeticiones por tratamiento. Entonces se usa un método de comparación de medias de celdas, por ejemplo el SNK. El error estándar para medias de celdas es 5.04 y los valores críticos de comparación W_p son:

p	2	3	4	5	6	7	8
$q_{0.05\,(p,16)}$	3.00	3.65	4.05	4.33	4.56	4.74	4.90
W_p	15.1	18.4	20.4	21.8	23.0	23.9	24.7

Las medias ordenadas en forma creciente son:

35.3 43.3 45.0 47.3 51.3 53.3 55.7 58.7.

Puesto que la máxima diferencia (58.7 – 35.3 = 23.4) es menor que 24.7, se concluye que todas las medias son similares y no se hacen otras comparaciones. La sensibilidad de esta prueba es inferior a la de la prueba F de la interacción, en efecto, aunque se concluyó una interacción débil, la prueba SNK no la detectó.

Otro análisis que puede ser interesante cuando hay interacción, es la comparación de medias de los niveles de un factor para un solo nivel del otro factor; por ejemplo, el insecticida para la raza R_2. De la tabla 9.3 se concluye que no hay diferencias significativas entre las concentraciones, sin embargo, se aplica el método SNK para comparar estas medias, teniendo en cuenta que las conclusiones solo se aplican al nivel R_2 del factor raza.

Medias ordenadas:	C_3	C_1	C_4	C_2
	35.3	43.3	47.3	58.7

La máxima diferencia es 58.7 – 35.3 = 23.4 > W_4 = 4.05 × 3.57 = 14.45. Luego, el porcentaje de mortalidad para la concentración C_2 es mayor que para la concentración C_3, en la raza R_2. La diferencia $C_4 - C_3$ = 12.0 es menor que W_3 = 3.65 × 3.57 = 13.03 por lo que el grupo $\{C_3, C_1, C_4\}$ se considera homogéneo.

La diferencia $C_2 - C_1$ = 15.4 > 13.03, también $C_2 - C_4$ = 11.4 > W_2 = 3.00 × 3.57 = 10.7. Se concluye que la concentración C_2 produce un porcentaje de mortalidad significativamente mayor que las demás concentraciones en la raza R_2.

En forma similar pueden compararse por separado las razas para cada concentración. Como solo se tienen dos razas, la diferencia entre razas será significativa si es mayor que W_2 = 3.00 × 2.52 = 7.56.

Cuando la interacción es significativa en un diseño bifactorial con un factor cuantitativo, el investigador puede estar interesado en estudiar la curvatura sobre cada uno de los niveles del factor cualitativo. Igualmente, si ambos factores son cuantitativos, puede estudiarse la no aditividad mediante polinomios ortogonales. Varios libros incluyen este tema y lo desarrollan ampliamente; aquí no se hará por razones de espacio, pero el lector puede consultar los textos de Anderson y McLean (1974), Hicks y Turner (1999) y Montgomery (2007).

9.3 Modelos mixtos

En la sección anterior se analizó el modelo I para un experimento factorial completamente aleatorizado. Los modelos aleatorios y mixtos tienen un análisis similar, pero teniendo en cuenta las respectivas componentes de varianza y sus valores esperados. En la tabla 9.5 se muestra el ANOVA para estos dos modelos; aquí se pueden reconocer las diferencias en las pruebas F para efectos principales e interacciones.

Considérese un experimento donde el factor A es fijo y el factor B es aleatorio (ejemplo 9.2). El modelo matemático correspondiente puede escribirse como:

$$y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \varepsilon_{ijk}. \tag{9.2}$$

con el mismo significado del modelo (9.1), excepto para los términos B_j y $(\alpha B)_{ij}$ que son aleatorios. El efecto B_j es una variable con distribución $N(0, \sigma^2_B)$, independiente de los errores ε_{ijk} y $(\alpha B)_{ij}$ es una variable normal con media cero y varianza $\sigma^2_{\alpha B}\,(a-1)/a$; por último, se supone que:

$$\sum_i \alpha_i = \sum_i (\alpha B)_{ij} = 0.$$

Este modelo se conoce como modelo mixto estándar, Montgomery (2007). El siguiente ejemplo ilustra el análisis para un modelo mixto.

Ejemplo 9.2 Los siguientes datos fueron obtenidos por el biólogo John Jairo Ramírez en un estudio realizado en el lago del Parque Norte de Medellín (Colombia). El objetivo propuesto fue analizar la variación de la temperatura del agua del lago a diferentes profundidades y durante las 24 horas del día. El factor tiempo puede considerarse aleatorio, aunque los muestreos se hayan hecho a intervalos más o menos iguales (véase tabla 9.6).

Los cálculos de las sumas de cuadrados son similares a los realizados para el ejemplo 9.1, y los respectivos *gl* se calculan como en la tabla 9.5. La tabla del ANOVA resultante es la tabla 9.7. Tanto la interacción como el factor *B* —aleatorio— se prueban con *MCE* como denominador, pero el factor *A* debe probarse con el cociente $F = MCA / MCAB$, de acuerdo con la tabla 9.5.

Como puede observarse, no hay interacción entre la profundidad y las horas del día ($p = 0.088$). La figura 9.2 muestra paralelismo en las líneas de temperatura promedio; en las horas de la mañana todas suben y en las horas de la tarde y de la noche todas bajan.

Cuando la interacción es no significativa se continúa el análisis con las pruebas de los efectos principales. Hay diferencias de temperatura entre las profundidades, ($p = 0.003$). Mediante el método de Tukey con $DHS = q_{0.05,5,40}\sqrt{0.27/16} = 0.053$, se tiene el resultado que se muestra en la figura 9.3.

Se concluye que hay diferencias de temperaturas entre la superficie y las profundidades mayores que 0.75 m, como también entre 0.50 y 1.50 m. Claro que el método detecta como significativas las diferencias mayores de medio grado centígrado.

Por otra parte, se concluye que la temperatura del agua es variable en el transcurso de las 24 horas. Una estimación de la varianza σ_B^2 se obtiene a partir de la *MCB*. Así:

$$\hat{\sigma}^2 + ar\hat{\sigma}_B^2 = MCB, \text{ luego } \hat{\sigma}_B^2 = \frac{1}{ar}(MCB - MCE).$$

Tabla 9.5 Tabla del ANOVA para los modelos factoriales aleatorio y mixto

Fuente de variación	*gl*	*MC*	*Modelo aleatorio*	*Modelo mixto*
Factor *A*	$a-1$	*MCA*	$\sigma^2 + r\sigma_{AB}^2 + br_{\sigma A}^2$	$\sigma^2 + r\sigma_{\alpha B}^2 + br\Sigma\alpha_i^2 / (a-1)$
Factor *B*	$b-1$	*MCB*	$\sigma^2 + r\sigma_{AB}^2 + ar_{\sigma B}^2$	$\sigma^2 + ar_{\alpha B}^2$
Interacción *AB*	$(a-1)(b-1)$	*MCAB*	$\sigma^2 + r\sigma_{AB}^2$	$\sigma^2 + r\sigma_{\alpha B}^2$
Error	$ab(r-1)$	*MCE*	σ^2	σ^2
Total	$abr-1$			

Tabla 9.6 Temperaturas del agua (°C) del lago del Parque Norte, Medellín, según el ejemplo 9.2

Profundidad (m)	*Tiempo (8 muestras durante 24 horas al día)*								*Medias*
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	
0.00	25.3	25.1	26.3	25.5	25.5	24.9	24.6	24.6	25.4
	26.0	27.8	26.5	25.6	25.6	24.9	24.4	24.0	
0.50	24.2	24.8	26.2	25.6	25.4	24.9	24.8	24.3	25.2
	24.8	27.0	26.3	25.6	25.5	25.1	24.2	23.9	
0.75	24.1	24.1	25.1	25.6	25.4	24.9	24.8	24.3	24.8
	24.0	25.5	25.8	25.6	25.4	25.0	24.1	23.7	
1.00	24.0	23.8	24.4	25.6	25.3	24.9	24.8	24.2	24.7
	23.8	24.6	25.4	25.3	25.4	25.0	24.9	23.6	
1.50	24.0	23.6	24.0	24.6	24.8	24.8	24.8	24.2	24.5
	23.8	24.6	24.7	25.1	25.3	25.0	24.8	23.6	
Medias	24.4	25.1	25.5	25.4	25.4	24.9	24.6	24.0	24.9

Tabla 9.7 Tabla del ANOVA para el ejemplo 9.2

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Profundidad (A)	8.87	4	2.22	2.22 / 0.43 = 5.16	0.003
Tiempo (B)	19.00	7	2.71	2.71 / 0.27 = 10.04	<0.001
Interacción (AB)	11.97	28	0.43	0.43 / 0.27 = 1.59	0.088
Error	10.97	40	0.27		
Total	50.81	79			

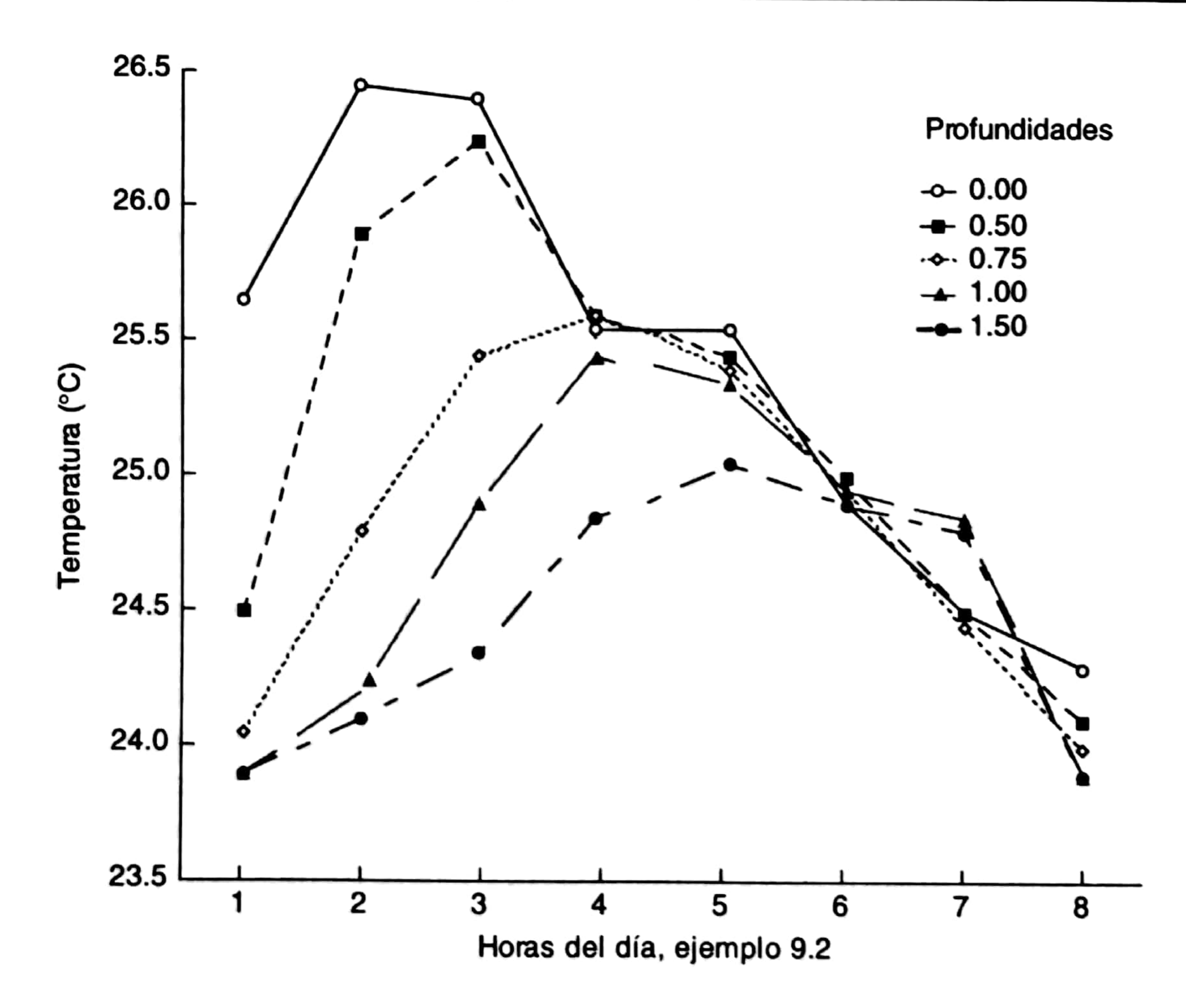


Figura 9.2 Interacción entre profundidades y tiempos para el ejemplo 9.2

$$\hat{\sigma}_B^2 = (2.71 - 0.27)/10 = 0.244.$$

La varianza de las lecturas de la temperatura es la suma de dos componentes: $\hat{\sigma}_y^2 = 0.244 + 0.27 = 0.514$.

Entonces, la hora del día contribuye con el 47.5% a la variación total. En otras palabras, alrededor de la mitad de la variación de la temperatura del agua es atribuible a la hora del día en que se realiza la medida.

Es procedente aclarar que la varianza total es la suma de tres componentes $\sigma_y^2 = \sigma_B^2 + \sigma_{\alpha B}^2 + \sigma^2$; no obstante la componente $\sigma_{\alpha B}^2$ de la interacción se supuso igual a cero porque se concluyó que su efecto era insignificante.

Profundidades: 0.00 0.50 0.75 1.00 1.50 m

Figura 9.3 Resultado del método de Tukey para el ejemplo 9.2

9.4 Experimentos bifactoriales con una observación por celda

Por motivos de costo o tiempo, en algunos experimentos factoriales solo se realiza una repetición por tratamiento. El problema de una observación por celda es que no es posible separar el efecto de interacción entre factores del error experimental, o sea, no es posible obtener medias cuadráticas separadas para cada fuente de variación; se dice, entonces, que los efectos de la interacción están confundidos con los del error experimental. Muchas veces, estudios previamente realizados permiten suponer aditividad entre los factores de modo que, al no tener en cuenta el efecto de interacción, la suma residual se asigna en su totalidad al error experimental. El modelo correspondiente es el mismo modelo (9.1) o (9.2) simplificado:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}; \text{ con } k = 1. \qquad (9.3)$$

Este modelo tiene la apariencia del modelo (7.1) para un diseño de bloques, sin embargo, representan diseños diferentes e, igualmente, su interpretación es distinta, aunque el ANOVA para los datos sea el mismo.

Los modelos (9.3) y (7.1) representan situaciones experimentales diferentes: en el modelo (9.3) los *ab* tratamientos del diseño factorial se ejecutan completamente al azar; mientras que en (7.1), los *k* tratamientos del factor se aleatorizan independientemente en cada bloque. Es decir, la manera como se obtienen los datos y su interpretación son totalmente diferentes para ambos diseños. Con el modelo (9.3) pueden realizarse las pruebas para efectos principales de cada factor en la forma usual. Si no se tiene historia sobre la interacción, y esta se considera importante, será necesario tomar dos o más observaciones por celda para poder evaluarla.

Cuando ambos factores son cualitativos, es posible investigar la no aditividad entre ellos usando la prueba de Tukey (véase sección 7.4). La suma de cuadrados de no aditividad con 1 *gl* da una indicación de la existencia o no de la interacción. En el caso afirmativo, puede ensayarse una transformación de los datos que remueva la interacción; esto ocurre cuando los datos son tomados en una escala que no es la apropiada. Si la interacción no desaparece mediante transformaciones, hay indicios de que los factores tienen comportamiento complejo que no es posible analizar con una única observación por celda.

De otro lado, en la descripción de la relación entre la variable de respuesta y los factores, los factores cuantitativos son generalmente más eficientes que los cualitativos. Se ha encontrado que la mayoría del efecto cuantitativo puede expresarse mediante una componente lineal con 1 *gl*. Por tanto, para obtener una estimación de la no aditividad, lo mejor es usar la interacción entre el factor cualitativo y el efecto lineal del factor cuantitativo.

El problema de examinar la no aditividad en un experimento bifactorial con una sola repetición, no es simple: si ambos factores son cualitativos, el mejor método es el de Tukey; si al menos un factor es cuantitativo, el procedimiento más recomendado es el de los polinomios ortogonales, tema que se encuentra tratado con detalle en Hicks y Turner (1999).

Ejemplo 9.3 Se diseñó un experimento para estudiar los efectos de cuatro dietas sobre el colesterol, medidos en el hígado de ratas. Se tuvo en cuenta también el factor peso con tres niveles: pesado, mediano y ligero. Los resultados, presentados en la tabla 9.8, fueron tomados de Gill (1978, vol. 2), y adaptados con el fin de darle un sentido más real al ejemplo.

En la tabla 9.9 se incluye, además del ANOVA general, la prueba de no aditividad de Tukey, la cual resulta no significativa; únicamente el efecto principal de las dietas es significativo. Se puede, entonces, hacer comparaciones entre las diferentes dietas usando un error estándar igual a $\sqrt{(MCE/a)} = \sqrt{0.0106/3} = 0.06$.

Por ejemplo, la comparación entre el control y las demás dietas puede analizarse mediante el contraste $Q = \mu_1 - (1/3)(\mu_2 + \mu_3 + \mu_4)$. Una estimación de Q es $\hat{Q} = 1.63 - (1/3)$ $(1.78 + 2.04 + 1.83) = -0.25$, y la suma de cuadrados es $SC(\hat{Q}) = r\hat{Q}^2 / \Sigma c_j^2 = 3 \times 0.025 / (4/3) = 0.14$. Entonces $F = 0.14/0.0106 = 13.2$ ($p = 0.015$). Se concluye que el efecto sobre el colesterol de la dieta control es inferior al efecto promedio de las otras dietas.

La comparación de medias por pares, usando el método de Tukey, se representa en la figura 9.4.

Desde luego que las dos pruebas no deben hacerse simultáneamente, hay que elegir una de ellas en lo posible, antes de realizar el experimento.

9.5 Experimentos bifactoriales en un DBA

Los tres diseños básicos, DCA, DBA y DCL, son utilizados tanto en experimentos unifactoriales como en los de dos o más factores. Las razones para usar bloques son las mismas que se mencionaron en el capítulo 7. Con la formación de bloques se busca no solo reducir el error experimental sino también resolver el problema de un número insuficiente de *ue* homogéneas, o la imposibilidad de realizar varias repeticiones del experimento en un mismo tiempo.

El modelo matemático apropiado para un experimento de dos factores con bloques aleatorios —modelo I— tiene la siguiente forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + D_k + \varepsilon_{ijk}, \qquad (9.4)$$

donde D_k representa el efecto de los bloques y se supone que tiene distribución $N(0, \sigma_D^2)$. Se supone además que no existe interacción entre bloques y tratamientos;

Tabla 9.8 Datos correspondientes al ejemplo 9.3

Peso	*Dietas*			
	Control	*Grasa vegetal*	*Grasa animal*	*Ambas grasas*
Pesado	1.60	1.97	2.07	1.85
Mediano	1.63	1.75	2.06	1.91
Ligero	1.67	1.63	2.00	1.72

Tabla 9.9 Tabla del ANOVA para los datos del ejemplo 9.3

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Peso	0.0290	2	0.0145	1.37	0.34
Dieta	0.2582	3	0.0860	8.11	0.023
No aditividad	0.0021	1	0.0021	0.20	0.67
Error	0.0529	5	0.0106		
Total	0.3422	11			

Control	Grasa vegetal	Ambas grasas	Grasa animal
1.63	1.78	1.83	2.04

Figura 9.4 Comparación de medias por pares para el ejemplo 9.3

en el caso en que exista, será pequeña y se considerará confundida con el error experimental. La tabla del ANOVA para este estudio es la 9.10.

Las medias cuadráticas esperadas indican que tanto los efectos principales como la interacción, pueden probarse usando como denominador la *MCE*

Ejemplo 9.4 Se encerró un cultivo de peces en jaulas y estas se colocaron en un lago, con el objeto de investigar el efecto de tres niveles de un ingrediente artificial, adicionado al alimento, sobre el crecimiento de tres especies de peces. Las especies se designaron como *x*, *y* y *z*, y los niveles del ingrediente como 0, 100 y 200 g/kg de alimento. Las nueve combinaciones de tratamientos entre especies e ingredientes se asignaron al azar a las jaulas de cultivo, formando tres conglomerados o bloques iguales (véase tabla 9.11). La variable de respuesta fue el aumento promedio mensual del peso (en gramos) durante tres meses (adaptado de Bazigos, 1976).

Las sumas de cuadrados se calculan para cada fuente de variación en la forma usual. Las especies de peces son el factor *A* y el ingrediente es el factor *B*. La letra D representa los bloques. Así:

$C = (413)^2/27 = 6317.37.$

$SCT = 35^2 + 7^2 + \cdots + 9^2 + 11^2 - C = 8593 - 6317.37 = 2275.63.$

$SCD = 1/9\,[(133)^2 + (146)^2 + (134)^2] - C = 6329 - C = 11.63.$

$SCA = 1/9\,[(150)^2 + (112)^2 + (151)^2] - C = 109.85.$

$SCB = 1/9\,[(239)^2 + (95)^2 + (79)^2] - C = 1725.63.$

$SCAB = 1/3\,[(92)^2 + (33)^2 + \cdots + (28)^2] - C - SCA - SCB = 134.82.$

$SCE = SCT - SCA - SCB - SCAB - SCD.$
$= 2275.63 - 109.85 - 1725.63 - 134.82 - 11.63 = 293.70.$

Como lo muestra la tabla 9.12, solamente los niveles del ingrediente influyen en forma diferente sobre el incremento de peso. La diferencia se observa claramente al relacionar el

Tabla 9.10 Tabla del ANOVA para un experimento bifactorial —modelo I— con bloques

Fuente de variación	*SC*	*gl*	*MC*	*MC esperada*
Bloques	*SCD*	$r-1$	*MCD*	$\sigma^2 + ab\,\sigma_D^2$
Factor A	*SCA*	$a-1$	*MCA*	$\sigma^2 + br\sum\alpha_i^2 / (a-1)$
Factor B	*SCB*	$b-1$	*MCB*	$\sigma^2 + ar\sum\beta_j^2 / (b-1)$
Interacción AB	*SCAB*	$(a-1)(b-1)$	*MCAB*	$\sigma^2 + r\sum(\alpha\beta)_{ij}^2 / (a-1)(b-1)$
Error	*SCE*	$(ab-1)(r-1)$	*MCE*	σ^2
Total	*SCT*	$abr-1$		

Tabla 9.11 Datos correspondientes al ejemplo 9.4

Especie	*Ingrediente*	*Bloques*			*Totales*			
		I	*II*	*III*	*Ingredientes*			*Especies*
	0	35	30	27	92			
X	100	7	8	18		33		
	200	14	4	7			25	150
	0	14	25	19	58			
Y	100	11	8	9		28		
	200	8	13	5			26	112
	0	27	33	29	89			
Z	100	10	15	9		34		
	200	7	10	11			28	151
Totales		133	146	134	239	95	79	413

promedio del nivel cero (26.56 g) con los otros dos promedios (10.56 y 8.78 g). Se concluye que la formación de bloques no fue exitosa en este experimento.

9.6 Experimentos factoriales 2^n

Entre los experimentos factoriales simétricos, el más sencillo de realizar es el factorial 2^n, el cual está compuesto de n factores cada uno con dos niveles. Estos experimentos permiten analizar un número grande de factores a la vez, con el objeto, por lo general, de tamizar o seleccionar aquellos que muestran mayor influencia en la variable de respuesta. Además, son experimentos apropiados para hacer estudios preliminares, cuyo fin sea reconocer los factores con efectos significativos que posteriormente serán incluidos en diseños más completos y complejos.

Los diseños para experimentos 2^n pueden estudiarse con los modelos corrientes del ANOVA. Sin embargo, se han ideado otras metodologías que resultan más simples, sobre todo si no se emplea *software* estadístico. Primero se verá el diseño 2^2, que es el más simple de todos.

Tabla 9.12 Tabla del ANOVA para los datos del ejemplo 9.4

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Bloques (D)	11.63	2	5.82	0.32	0.73
Especies A	109.85	2	54.93	2.99	0.079
Ingredientes B	1 725.63	2	862.82	46.99	<0.001
Interacción AB	134.82	4	33.71	1.84	0.17
Error	293.70	16	18.36		
Total	2 275.63	26			

Considérense dos factores A y B con niveles a_0, a_1 y b_0, b_1, respectivamente. Las combinaciones de tratamientos son a_0b_0, a_1b_0, a_0b_1, a_1b_1. Es usual representar los niveles en la siguiente forma:

a_0: nivel bajo de A.

a_1: nivel alto de A.

b_0: nivel bajo de B.

b_1: nivel alto de B.

Los tratamientos o productos de niveles tienen diferentes nomenclaturas numéricas y literales; aquí se utilizará una de las más comunes:

$(1) = a_0b_0$, indica que A y B están en el nivel bajo.

$a = a_1b_0$, indica que A está en el nivel alto y B en el nivel bajo.

$b = a_0b_1$, indica que B está en el nivel alto y A en el nivel bajo.

$ab = a_1b_1$, indica que A y B están en el nivel alto.

Los resultados del experimento se simbolizan en una tabla de doble entrada que tiene la apariencia de la tabla 9.13.

Ejemplo 9.5 Para estudiar el efecto de agregar vitamina B_{12} conjuntamente con un antibiótico a un alimento para cerdos, se diseñó un experimento factorial 2×2 cuya respuesta fue el aumento promedio diario de peso de los cerdos. Los factores y sus niveles se muestran en la tabla 9.14.

La tabla de resultados, dados en g, adaptada de Snedecor y Cochran (1984), es la tabla 9.15 con tres repeticiones por tratamiento.

Para efectuar el análisis de estos datos, se calculan los promedios por celda y se forma la tabla 9.16 con las siguientes definiciones.

Efectos simples. Se llama efecto simple a la diferencia entre medias del nivel alto y del nivel bajo de un factor, dejando fijo el nivel del otro factor.

Por ejemplo, 14 es el efecto simple del factor B en el nivel bajo de A y 231 es el efecto simple de B en el nivel alto de A.

Efectos principales. El efecto principal de un factor es el promedio de los efectos simples del factor.

Efecto principal del factor $B = (1/2)(14 + 231) = 122.5$.

Tabla 9.13 Tabla de doble entrada para experimentos factoriales 2^2

Factor A	*Factor B*	
	b_0	b_1
a_0	(1)	b
a_1	a	ab

Tabla 9.14 Factores y niveles para el ejemplo 9.5

Factores	*Nivel bajo*	*Nivel alto*
A: antibiótico	$a_0 = 0.00$ mg	$a_1 = 0.40$ mg
B: vitamina B_{12}	$b_0 = 0.00$ mg	$b_1 = 0.50$ mg

$$\textit{Efecto principal del factor } A = (1/2)(-71 + 146) = 37.5.$$

Interacción. La interacción entre los factores A y B es el promedio de la diferencia de los efectos simples de uno de los dos factores.

$$\textit{Interacción } AB = (1/2)(231 - 14) = 108.5,$$

o también

$$\textit{Interacción } AB = (1/2)(146 + 71) = 108.5.$$

Si los factores A y B no interaccionan, o sea, no tienen influencia conjunta sobre la variable de respuesta, los efectos simples de cada factor son similares entre sí, y su diferencia será muy pequeña.

También pueden usarse los totales de celda en vez de los promedios; en este caso, los símbolos (1), a, b, ab representan las sumas de las r repeticiones de cada celda, respectivamente. Entonces, debe diferenciarse entre efecto promedio y efecto total de un factor, dependiendo de la cantidad —la media o el total— que aparezca en la celda. Aquí se seguirán usando promedios; sin embargo, en muchos textos utilizan totales de celda en vez de medias.

En forma literal, el efecto promedio del factor B se define como el promedio de los efectos simples representados por $b - (1)$ y $ab - a$. Entonces:

$$\textit{Efecto promedio de } B = (1/2)[b - (1) + ab - a] = 1/2\,[ab + b - a - (1)].$$

Similarmente, se definen los otros efectos, así:

$$\textit{Efecto promedio de } A = (1/2)\,[ab + a - b - (1)].$$

$$\textit{Efecto promedio de } AB = (1/2)\,[ab - a - b + (1)].$$

Tabla 9.15 Datos correspondientes al ejemplo 9.5

Antibiótico	*Vitamina B_{12}* 0.0	*Vitamina B_{12}* 0.5	*Totales*
0.0	590	572	
	540	549	3281
	490	540	
0.4	476	690	
	454	708	3507
	476	703	
Totales	3026	3762	6788

Tabla 9.16 Tabla de promedios de tratamientos para el ejemplo 9.5

Antibiótico	*Vitamina B_{12}* 0.0	*Vitamina B_{12}* 0.5	*Diferencia*
0.0	540	554	14
0.4	469	700	231
Diferencia	−71	146	

Como puede verse, el efecto promedio es un contraste entre medias de celdas; además, los contrastes A, B y AB son ortogonales entre sí.

Ahora bien, hay tres formas de estimar los efectos de los factores:

1) usando la definición,
2) mediante una tabla de signos y
3) por el algoritmo de Yates.

Para el diseño 2^2, la tabla de signos es la tabla 9.17, obtenida de los signos de los respectivos contrastes. Cuando el número de factores es grande, la tabla de signos se construye siguiendo algunas normas simples, Cochran y Cox (1995).

El efecto principal se calcula como la suma de las respuestas, cada una de ellas con el signo que le corresponde en la respectiva columna, dividida por 2^{n-1}. Por ejemplo, el efecto principal de A es:

$$A = (1/2)(-540 + 469 - 554 + 700) = 37.5.$$

La columna encabezada por I solo tiene signos positivos; al dividir la suma por 2^n, se obtiene la media general.

El algoritmo de Yates permite hallar los efectos de los factores mediante un procedimiento iterativo: a partir de la columna

Tabla 9.17 Tabla de signos para el diseño 2^2

Tratamientos	*Efectos*				*Respuesta*
	I	*A*	*B*	*AB*	$\bar{y}_{ij}$
(1)	+	–	–	+	540
a	+	+	–	–	469
b	+	–	+	–	554
ab	+	+	+	+	700

Tabla 9.18 Algoritmo de Yates para un diseño 2^2

Tratamientos	*Respuestas*	*(1a)*	*(2a)*	*Efectos*
(1)	540	1009	2263	$565.7 = \bar{y}$
a	469	1254	75	$37.5 = A$
b	554	–71	245	$122.5 = B$
ab	700	146	217	$108.5 = AB$

de respuestas promedio se construyen otras dos columnas, o tantas columnas como factores se tengan (véase tabla 9.18).

La mitad de los elementos de la primera columna resultan de sumar los valores adyacentes por pares (540 + 469 y 554 + 700); la otra mitad de restar los mismos valores en el siguiente orden: el segundo menos el primero y el cuarto menos el tercero (469 – 540 y 700 – 554).

La segunda columna se obtiene a partir de la primera repitiendo el mismo proceso. Los efectos *A*, *B* y *AB*, se obtienen al dividir la última columna por 2^{n-1}, excepto el primer valor que se divide por 2^n para obtener la media general.

El efecto principal de *A* es 37.5 y significa que si se pasa del nivel bajo de *A* al nivel alto, se produce incremento promedio diario de peso de 37.5 g. Es decir, el animal aumenta más rápidamente de peso cuando se le suministra antibiótico que cuando no se le suministra. Sin embargo, para saber qué tan significativo es este aumento de peso, y dado que el valor 37.5 es un número absoluto y no puede juzgarse como pequeño o grande, es necesario compararlo con un patrón de referencia. Generalmente este patrón no se conoce, no existe o no es necesario, ya que es posible realizar una prueba de significación estadística para estos efectos.

Tanto los efectos principales como las interacciones son contrastes de medias, y su significancia puede probarse mediante los métodos del capítulo 6.

Las sumas de cuadrados se calculan de la siguiente manera:

$$SC(efecto) = 2^{n-2}\, r\, (\text{efecto})^2.$$

$$SC(A) = 3(37.5)^2 = 42\,18.8.$$

$$SC(B) = 3(122.5)^2 = 45018.8.$$

$$SC(AB) = 3(108.5)^2 = 35316.8.$$

Cada una de estas sumas de cuadrados tiene 1 *gl*. Para calcular la *SCE* se sigue el procedimiento del ANOVA, o sea, se calcula la *SCT* y de esta se restan las otras sumas. Entonces:

$$SCT = (590)^2 \ldots + (700)^2 - (6788)^2 / 12 = 91080.7.$$

$$SCE = 91080.7 - (4218.8 + 45018.8 + 35316.8) = 6526.6.$$

$$MCE = \frac{SCE}{2^n\,(r-1)} = 6526.6/8 = 815.8.$$

La varianza del error experimental depende del número de repeticiones; de modo que si no se repite el experimento, la varianza no es estimable y no pueden realizarse las pruebas estadísticas.

Para analizar el experimento, primero se prueba la interacción *AB* mediante el cociente $F = MCAB\,/\,MCE = 35316.8/815.8 = 43.3$. Como $p < 0.001$, la interacción es altamente significativa, por lo que el análisis deberá centrarse en este efecto y no en los

efectos aislados de cada factor. Teniendo como referencia la tabla 9.16, se observa que, en ausencia de antibiótico, la vitamina es poco efectiva —aumenta de 540 a 554 g en promedio—; mientras que en presencia del antibiótico, el crecimiento es mucho mayor —pasa de 469 a 700 g en promedio—. Una explicación posible, según los autores, es que el antibiótico controla la flora intestinal indeseable; de modo que sin el antibiótico, la flora intestinal neutraliza la acción de la vitamina.

9.7 Experimentos factoriales 2^3

Un experimento con tres factores tendrá ocho tratamientos, los cuales se escriben en un orden fijo llamado orden estándar. Sean tres factores *A, B* y *C,* cada uno con dos niveles llamados bajo y alto. De acuerdo con la nomenclatura anterior para los diseños 2^2, el orden estándar es como sigue: (1), *a, b, ab, c, ac, bc, abc*. Ahora bien, los efectos simples de un factor se determinan por diferencias entre combinaciones con el nivel y sin el nivel del factor. Por ejemplo, el efecto principal de *A* es el promedio de los cuatro efectos simples de dicho factor:

$$A = (1/4)[(abc - bc) + (ab - b) + (ac - c) + (a - (1)].$$

En forma análoga se definen los otros efectos. En total se tienen siete efectos y la forma más expedita de estimarlos es mediante el algoritmo de Yates, en vez de usar definiciones o la tabla de signos.

En este experimento se tienen tres interacciones de primer orden *(AB, AC* y *BC)* y una interacción de segundo orden *(ABC).* Esta última es más compleja de interpretar que las de primer orden, pero en la práctica casi siempre es pequeña, y por eso no se tiene en cuenta para un análisis simultáneo de los factores. En experimentos con un mayor número de factores, las interacciones de segundo orden y de orden superior se confunden con el error experimental y no se analizan independientemente.

Ejemplo 9.6 Se realizó un experimento de laboratorio para estudiar la influencia del fósforo, el nitrógeno y el carbono en la regulación del crecimiento del alga *Microcystis aeruginosa*. La respuesta fue la población algal medida espectrofotométricamente en unidades de absorbancia. Se hicieron dos repeticiones del experimento completamente al azar (véanse tablas 9.19 y 9.20).

Las estimaciones de los efectos se obtienen en la tabla 9.21. Como hay tres factores, se construyen tres columnas, según el procedimiento sugerido por Yates, y los efectos se obtienen dividiendo por $2^{n-1} = 4$ cada elemento de la última columna, excepto el primero. Por ejemplo, $A = -0.1310 / 4 = -0.033$.

Las sumas de cuadrados se obtienen elevando al cuadrado el efecto y multiplicando por $r(2^{n-2}) = 4$. Así, $SC(B) = 4(0.013)^2 = 0.0007$.

Además,

$$SCT = [(0.312)^2 + (0.576)^2 + \cdots + (0.814)^2] - (1 / 16)(7.57)^2 = 0.4391.$$

$$SCE = 0.4391 - 0.2473 = 0.1918.$$

$$MCE = (1 / 8)(0.1918) = 0.024.$$

La única suma de cuadrados mayor que *MCE* es la $SC(C) = 0.2153$. El cociente $F = 0.2153/0.024 = 8.97$ indica que el efecto principal de *C* es significativo ($p = 0.017$):

Tabla 9.19 Factores y niveles para el ejemplo 9.6

Factores	*Nivel bajo*	*Nivel alto*
A = fósforo (mg/L fosfato)	0.06	0.30
B = nitrógeno (mg/L nitrato)	2.00	10.00
C = carbono (mg/L bicarbonato)	20.00	100.00

Tabla 9.20 Datos correspondientes al ejemplo 9.6

Factor A	*Factor B*	*Factor C*	
		c_0	c_1
a_0	b_0	0.312	0.479
		0576	0.656
	b_1	0.412	0.465
		0.280	0.736
a_1	b_0	0.391	0.481
		0.309	0.631
	b_1	0.376	0.451
		0.201	0.814

solamente el carbono tiene influencia como regulador del crecimiento de esta especie de alga.

Los experimentos con un número grande de factores exigen recursos que muchas veces son difíciles de obtener. Un experimento 2^6 con dos repeticiones, exigiría 128 *ue* en dos grupos homogéneos, que no siempre se consiguen. Sin embargo, no es necesario realizar el experimento completo para obtener la información deseada. Cuando el conocimiento del fenómeno en estudio permite suponer que las interacciones más altas son insignificantes, o si el investigador no tiene interés en ellas, es posible obtener información sobre los efectos principales e interacciones de orden inferior, realizando solo una fracción del experimento. Estos procedimientos experimentales se llaman *diseños factoriales fraccionados*.

Otras técnicas experimentales que permiten reducir el tamaño de los diseños factoriales, sacrificando parte de la información pero ganando simplicidad y precisión, son la confusión de efectos con bloques y la subdivisión de parcelas, llamadas respectivamente, *diseños de confusión y de parcelas divididas,* Cochran y Cox (1995) y Montgomery (2007).

Ejercicios

9.1 Los siguientes resultados representan la velocidad de consumo de oxígeno de dos especies de lapas: *A. scabra (A. s)* y *A. digitalis (A. d),* a dos concentraciones salinas de agua de mar. Se hicieron dos repeticiones por tratamiento.

Tabla 9.21 Algoritmo de Yates para los datos del ejemplo 9.6

Tratamiento	*Respuesta* $\bar{y}_{ij.}$	*1ª*	*2ª*	*3ª*	*Efectos*	*SC*
(1)	0.4440	0.7940	1.4285	3.7850	0.473 = $\bar{y}$	
a	0.3500	0.6345	2.3565	–0.1310	–0.033 = A	0.0043
b	0.3460	1.1235	–0.1515	–0.0500	–0.013 = B	0.0007
ab	0.2885	1.2330	0.0205	0.0800	0.020 = AB	0.0016
c	0.5675	–0.0940	–0.1595	0.9280	0.232 = C	0.2153
ac	0.5560	–0.0575	0.1095	0.1720	0.043 = AC	0.0074
ac	0.6005	–0.0115	0.0365	0.2690	0.067 = BC	0.0180
abc	0.6325	0.0320	0.0435	0.0070	0.002 = ABC	0.00001

Concentración	*Especies*	
	A. s.	*A. d.*
100%	13.6	10.4
	8.9	5.5
50%	18.8	13.4
	9.7	14.5

a. Calcular los efectos principales y la interacción entre factores.
b. Realizar el análisis de significación según las hipótesis planteadas.
c. Construir un gráfico de totales que muestre el comportamiento de los factores en estudio.

9.2 Con el fin de comparar los rendimientos proteínicos de cuatro variedades de avena, Gaviria y otros (1989) diseñaron un experimento 3 × 4 con dos repeticiones. El factor A representa el estado fisiológico en el momento del corte y el factor B la variedad de la semilla. Los resultados, en porcentajes promedios y en base seca, fueron:

Estados	*Variedad de la avena*			
	Bacatá	*Cajicá*	*Cayuse*	*Nehuen*
Prefloración	9.55	11.76	7.16	10.68
	10.32	11.60	7.23	11.26
Floración	8.61	11.05	6.92	6.96
	8.06	9.80	6.69	10.80
Grano lechoso	6.77	7.26	6.68	9.92
	6.02	8.72	4.53	13.58

a. Construir la tabla del ANOVA y verificar los supuestos.
b. Obtener las conclusiones de las pruebas F.
c. Graficar medias o totales de celda y corroborar las conclusiones obtenidas en *b.*
d. Estimar las medias de variedades y de estados. Interpretar los resultados.
e. Comparar los porcentajes promedio de cada variedad en cada uno de los tres estados fisiológicos.

9.3 Se midió el contenido de calcio en la sangre (en mg/100 mL) de peces sometidos a uno de tres tratamientos hormonales (Zar, 1999).

Sexo	*Tratamientos*		
	T_1	T_2	T_3
Machos	16.87	19.07	32.45
	16.18	18.77	28.71
	17.12	17.63	34.65
	16.83	16.99	28.79
	17.19	18.04	24.46
Hembras	15.86	17.20	30.54
	14.92	17.64	32.41
	15.63	17.89	28.97
	15.24	16.78	28.46
	14.80	16.92	29.65

a. Construir la tabla del ANOVA y chequear los supuestos.
b. Construir la tabla de medias de celdas y de errores estándares. Graficar las medias para los tratamientos.
c. Comparar los tratamientos hormonales.
d. Calcular un intervalo de confianza del 95% para la media del tratamiento T_3.

9.4 Se hizo un estudio en ratones para comparar la efectividad relativa de cuatro vacunas antivirales junto con seis cantidades diferentes de un adyuvante, que fue diseñado para aumentar la producción de anticuerpos. Para cada combinación vacuna-adyuvante se hicieron tres repeticiones tomando suero de cada uno de tres ratones. La respuesta medida es la máxima numeración del tubo en una serie de diluciones en las que ocurrió la neutralización de la acción patogénica del virus (Gill, 1978).

Vacunas	*Adyuvante*					
	1	*2*	*3*	*4*	*5*	*6*
	5	2	3	7	3	7
V_1	6	4	3	4	8	8
	5	4	6	3	6	3
	3	3	5	2	6	4
V_2	2	6	7	7	3	7
	4	3	6	4	4	6
	5	5	6	5	9	3
V_3	2	3	7	6	7	6
	2	6	4	7	4	8
	2	4	2	7	5	5
V_4	4	2	2	2	6	2
	2	3	2	3	2	4

a. Construir la tabla del ANOVA y verificar los supuestos.
b. Comparar los resultados promedio de las vacunas.
c. Observar la característica llamativa de las medias de los adyuvantes.

9.5 Se estimó el valor nutritivo de cierto fruto comestible, cultivado en diferentes regiones geográficas, y se analizaron tres frutos de cada una de cuatro variedades, en tres regiones seleccionadas al azar de un conjunto mayor (Daniel, 2004).

Región	*Variedades*			
	V_1	V_2	V_3	V_4
	6.9	11.0	13.1	13.4
A	6.2	7.3	9.9	13.5
	9.2	7.9	11.3	12.3
	9.2	5.1	7.1	13.1
B	7.7	9.4	13.7	8.6
	5.7	5.7	7.5	9.9
	5.2	9.3	8.1	10.8
C	5.5	6.6	10.6	12.3
	7.3	10.8	10.5	14.0

a. Considerar un modelo mixto, establecer el modelo matemático y sus supuestos.
b. Construir la tabla del ANOVA y obtener las conclusiones respectivas.
c. Estimar la variabilidad entre regiones.

9.6 En tres especies de cítricos se determinó la razón de la superficie foliar a peso seco para tres condiciones de sombra (Snedecor y Cochran, 1984).

Condiciones de sombra	*Naranjo*	*Toronja*	*Mandarina*
Sol	112	90	123
Media sombra	86	73	89
Sombra	80	62	81

a. Construir la tabla del ANOVA, incluyendo la prueba de no aditividad de Tukey.
b. Analizar los efectos principales de cada factor en estudio.

9.7 Se diseñó un experimento para comparar tres variedades de trigo y su repuesta a la aplicación de sulfato de amonio a cuatro niveles: 0, 10, 20 y 30 kg por hectárea. El experimento se realizó en cuatro bloques al azar y el rendimiento obtenido (kg por 100 m^2) se da en la tabla siguiente (adaptado de Panse y Sukhatme, 1963):

a. Construir la tabla del ANOVA y chequear los supuestos que hacen válido este análisis.
b. Discutir la efectividad en la formación de bloques.
c. Construir la tabla de medias y de errores estándares.
d. Efectuar las comparaciones de medias que tengan sentido en este experimento.

Variedades	*Sulfato de amonio*	*Bloques*			
		I	*II*	*III*	*IV*
V_1	0	7.6	6.6	8.2	7.3
	10	8.7	8.0	8.0	7.6
	20	7.4	6.3	8.5	8.6
	30	7.9	7.0	8.6	6.7
V_2	0	6.7	6.5	8.6	7.7
	10	8.5	8.0	10.2	8.0
	20	9.3	7.6	10.2	8.5
	30	8.5	9.0	9.5	8.9
V_3	0	7.2	8.2	7.7	7.7
	10	8.8	8.1	10.0	8.3
	20	9.9	7.8	10.8	10.2
	30	9.6	7.3	9.7	9.1

9.8 Orozco, H. O., citado por Gómez (1989), estudió en laboratorio el efecto de la inoculación del suelo con micorrizas y Rhizobium sobre el desarrollo de plantas de *Leucaena leucocephala.* Los siguientes datos son valores de las alturas (en cm) de las plantas, tomando el promedio de dos plantas por unidad experimental después de 90 días de la siembra.

Rhizobium	*Micorrizas*	
	Sin inóculo	*Con inóculo*
Sin inóculo	33.0	40.0
	33.5	39.5
	38.5	35.0
	33.5	32.0
Con inóculo	40.0	38.5
	43.5	30.0
	42.5	36.0
	40.5	39.0

a. Analizar estos datos usando el algoritmo de Yates.
b. Hallar la significancia de la interacción entre factores.
c. Estudiar los efectos simples, o sea el efecto del Rhizobium en ausencia y en presencia de la micorriza.
d. Comparar los niveles de micorriza para cada nivel de Rhizobium usando el método de Tukey. Establecer conclusiones.
e. Construir el gráfico de interacción entre factores.

9.9 Arenas y otros (1985) seleccionaron los siguientes factores, a dos niveles, para determinar las condiciones óptimas de extracción de peptinas utilizando la corteza de naranja como materia prima.

Factores	*Descripción*	*Nivel bajo*	*Nivel alto*
A	Temperatura	90 °C	96 °C
B	Tiempo	60 min	80 min
C	pH	1.5	2.5

La característica medida fue el grado de gelificación definido como el número de gramos de azúcar que puede combinarse con un gramo de pectina para dar un gel estándar. Los resultados de dos repeticiones se muestran en la tabla siguiente.

Temperatura (°C)	*Tiempo (min)*	*pH*	
		1.5	*2.5*
90	60	117.6	111.0
		118.4	109.0
	80	110.9	100.0
		111.1	106.0

Temperatura (°C)	*Tiempo (min)*	*pH*	
		1.5	*2.5*
96	60	124.0	114.0
		120.0	110.0
	80	111.0	113.0
		119.0	115.0

a. Usar el algoritmo de Yates para estimar los efectos de los factores.
b. Efectuar pruebas de hipótesis para efectos principales e interacciones.
c. Describir la influencia de cada uno de los factores en la variable de respuesta.

9.10 Con el fin de analizar el cambio de temperatura (en °C) de diferentes tipos de carne empacada en shopak, y con permanencia de cuatro días en refrigeración, Gómez y González (1991) obtuvieron los resultados que se muestran a continuación.

Tiempo	*Tocino*	*Costilla*	*Milanesa de cerdo*	*Posta*	*Milanesa de res*
Día 1	−1.0	3.0	0.2	2.0	2.0
	0.2	4.2	−1.2	0.3	3.5
	6.2	−1.0	0.0	3.7	3.3
Día 2	4.0	3.4	0.0	3.5	2.8
	3.5	5.4	2.4	2.8	4.7
	3.7	−3.7	−4.2	1.8	−3.5
Día 3	2.5	4.3	−0.5	2.0	2.7
	1.0	6.8	2.5	0.6	4.0
	1.2	−2.0	−4.3	2.0	−2.0
Día 4	2.5	2.3	−0.7	1.8	1.8
	1.3	4.2	−1.0	0.3	5.0
	5.6	−1.2	−2.0	4.8	0.0

a. Realizar el ANOVA correspondiente a estos datos y verificar los supuestos.
b. Establecer las diferencias entre las carnes en cuanto a cambios de temperatura.
c. Construir la tabla de cambios promedio de temperatura para las diferentes carnes.
d. Construir un gráfico de interacciones entre carnes y tiempo.

9.11 Aubad (1989) investigó el comportamiento del ritmo circadiano de peces sometidos a la influencia de varios metales pesados, disueltos en agua de consumo doméstico, con el fin de considerar este comportamiento como una alarma biológica en el control de aguas contaminadas. La variable medida fue el potencial bioeléctrico, registrado en un computador y obtenido como una varianza, de cada pez *(Carassius auratus)*.

Tiempo	*AD*	*Mercurio*	*Cobre*	*Cromo*	*Plomo*
10:00	0.93	1.11	0.35	0.87	0.38
	0.87	1.14	0.27	0.29	0.46
	0.47	0.50	0.27	0.54	0.62
	1.36	0.90	0.28	1.16	0.54
10:30	0.28	0.78	0.26	1.51	0.42
	0.64	3.60	0.29	0.28	0.47
	0.45	0.50	0.31	0.63	0.66
	1.42	0.97	0.27	0.39	0.55
11:00	0.34	0.95	0.26	0.66	0.41
	0.95	1.33	0.28	0.29	0.32
	0.40	1.17	0.30	1.74	0.44
	1.23	1.24	0.28	0.23	0.37
11:30	0.32	0.70	0.30	0.31	0.38
	0.83	1.27	0.31	0.27	0.34
	0.24	1.83	0.29	0.52	0.37
	1.38	0.84	0.23	0.21	0.30

De una tabla más extensa se seleccionó la siguiente información: potencial bioeléctrico promedio de peces expuestos a agua de dilución, AD, y a las concentraciones máximas de mercurio, cobre, cromo y plomo permitidas por la OMS (1972). Las lecturas se hicieron durante dos horas.

a. Chequear los supuestos de homoscedasticidad y de normalidad.
b. Ensayar una transformación de datos. Sugerencia: $1/\sqrt{y}$.
c. Hacer comparaciones entre metales.

9.12 Trujillo y Pulido (1986) investigaron el efecto de las aguas de lavado de residuos sólidos de la industria minera, sobre la calidad de una corriente de agua receptora en el río Nechí en El Bagre, departamento de Antioquia (Colombia). Se utilizó un diseño bifactorial para analizar varios parámetros, entre estos los sólidos totales disueltos (mg/L) de dos tipos de muestra y tres tamaños de partículas.

Tipo de muestra	*Tamaño de partículas*		
	1600/1000	*1000/500*	*500/250*
Draga	32.0	34.0	38.0
	36.0	38.0	66.0
	34.0	54.0	54.0
Acarreo y motobomba	28.0	20.0	40.0
	28.0	38.0	34.0
	22.0	28.0	34.0

a. Construir y analizar la tabla del ANOVA para estos datos.
b. Construir el gráfico de interacciones.

10

Experimentos anidados o jerárquicos

10.1 Modelo mixto

Los experimentos factoriales analizados en el capítulo anterior se llaman experimentos de clasificación cruzada, porque en el respectivo diseño aparecen todos los niveles de un factor combinados con los niveles de los otros factores. Sin embargo, en algunos experimentos los niveles de un factor están anidados en los niveles de otro factor, de modo que no es posible cruzarlos para formar combinaciones de tratamientos. Esta situación se presenta cuando los niveles de un factor se originan por submuestreo de los niveles de otro factor.

Supóngase que se desea estudiar un rasgo particular del hombre colombiano. Para ello se eligen algunos departamentos o regiones al azar; de cada uno de ellos se toman municipios, también al azar; de cada municipio se muestrean hogares y, finalmente, se muestrean personas dentro de esos hogares. Este procedimiento es un submuestreo en varias etapas y su equivalente en la experimentación es el diseño anidado. Como no existen los mismos municipios en todos los departamentos, es imposible obtener interacción entre el factor municipio y el factor departamento: se dice que el factor municipio está anidado, o se submuestrea, en el factor departamento. Por otro lado, el diseño anidado puede analizarse mediante un modelo mixto o un modelo de componentes de varianza. El modelo I no es apropiado, ya que en el submuestreo los niveles son seleccionados al azar.

Considérese, primero, el modelo mixto para un diseño anidado con un factor A fijo y un factor B anidado en A. El modelo matemático tiene la forma:

$y_{ijk} = \mu + \tau_i + B_{(i)j} + \varepsilon_{(ij)k}$; con:

$i = 1, 2, \ldots, a$ tratamientos,

$j = 1, 2, \ldots, b$ unidades experimentales,

$k = 1, 2, \ldots, r$ repeticiones, (10.1)

donde y_{ijk} es la k-ésima observación en el subgrupo j-ésimo del grupo i-ésimo,

μ es la media poblacional, τ_i, es el efecto fijo del nivel i-ésimo del factor A, $B_{(i)j}$ es el efecto aleatorio del j-ésimo subgrupo dentro del i-ésimo grupo y $\varepsilon_{(ij)k}$ es el error aleatorio de la repetición k-ésima en el subgrupo j-ésimo del i-ésimo grupo, el cual representa el error de submuestreo.

Se supone que $\sum_i \tau_i = 0$, que los $B_{(i)j}$ son independientes con distribución $N(0,\sigma^2)$ y que los $\varepsilon_{(ij)k}$ son independientes entre sí y con los $B_{(i)j}$, además, que estos errores tienen distribución $N(0,\sigma_s^2)$.

La tabla correspondiente del ANOVA para este diseño es la 10.1.

Ejemplo 10.1. Se diseñó un experimento de invernadero para observar el crecimiento de los tallos de plantas de menta cultivadas en una solución nutritiva y sometidas a tres periodos diferentes de luz diurna: 8, 12 y 16 horas diarias. Para cada periodo se seleccionaron tres materas y se midieron cuatro plantas muestreadas de cada matera. Los datos obtenidos se dan en la tabla 10.2 y el ANOVA correspondiente en la 10.3 (adaptado de Steel y Torne, 1997).

El factor A es el tiempo de luz diurna con tres niveles preseleccionados, factor fijo. El factor B es el conjunto de materas; estas representan las unidades experimentales y se supone que fueron seleccionadas al azar de un grupo grande de materas.

Al muestrear plantas de cada matera, se introduce otra fuente de variación llamada error de muestreo, de modo que el resultado de la medición de cada planta encierra dos fuentes de variación: una causada por la selección aleatoria de las plantas dentro de las materas, y la otra por la selección aleatoria de las materas. La primera está representada por la varianza σ_s^2; y la segunda por la varianza σ^2.

Con este diseño es posible probar dos hipótesis:

1. $H_0 : \sigma^2 = 0$ contra $H_1 : \sigma^2 > 0$.

La prueba se realiza con el cociente $F = MCE / MCS$, como lo sugiere la tabla 10.1. Si se rechaza H_0, se concluye que hay variación significativa entre las materas dentro del mismo tratamiento. La variación entre una planta y otra —error de muestreo— está incluida en la variación de matera a matera para el mismo periodo horario, puesto que las diferentes materas tienen a la vez diferentes plantas.

Se observa que este experimento no puede diseñarse como un factorial, porque en la misma matera no es posible someter las plantas a tres periodos de luz diferentes; es decir, tomar cuatro para exponerlas a ocho horas de luz, otras cuatro a doce horas de luz y otras cuatro a dieciséis horas de luz, diariamente.

Tabla 10.1 Tabla del ANOVA para un diseño anidado, modelo mixto

Fuente de variación	*SC*	*gl*	*MC*	*MC Esperada*	*F*
Entre grupos (factor A)	*SCA*	$a-1$	*MCA*	$\sigma_s^2 + r\sigma^2 + \frac{rb}{a-1}\Sigma\tau_i^2$	$\frac{MCA}{MCE}$
Entre subgrupos dentro de grupos (error experimental)	*SCE*	$a(b-1)$	*MCE*	$\sigma_s^2 + r\sigma^2$	$\frac{MCE}{MCS}$
Dentro de subgrupos (error de muestreo)	*SCS*	$ab(r-1)$	*MCS*	σ_s^2	
Total	*SCT*	$abr-1$			

Tabla 10.2 Crecimiento en una semana de los tallos de plantas de menta (datos en cm)

Horas de luz diurna	8			12			16		
Materas	1	2	3	4	5	6	7	8	9
Plantas (repeticiones)	8.5	6.5	7.0	6.0	6.0	6.5	7.0	6.0	11.0
	6.0	7.0	7.0	5.5	8.5	6.5	9.0	7.0	7.0
	9.0	8.0	7.0	3.5	4.5	8.5	8.5	7.0	9.0
	8.5	6.5	7.0	7.0	7.5	7.5	8.5	7.0	8.0
Totales de matera	32.0	28.0	28.0	22.0	26.5	29.0	33.0	27.0	35.0
Totales de horas	88.0			77.5			95.0		

2. H_0: $\tau_i = 0$ contra H_1: $\tau_i \neq 0$.

Esta hipótesis prueba igualdad de promedios de crecimiento por tratamiento. Si el cociente $F = MCA / MCE$ es mayor que el valor tabulado, se concluye que hay diferencias significativas entre los diferentes periodos de luz diurna.

Las sumas de cuadrados para el análisis de los datos se calculan así:

$$C = \frac{1}{abr}\left[\sum_i \sum_j \sum_k y_{ijk}\right]^2 = \frac{1}{36}(260.5)^2 = 1\,885.00.$$

$$SCT = \sum_i \sum_j \sum_k y_{ijk}^2 - C = 1\,952.75 - 1\,885.00 = 67.75.$$

$$SCA = \frac{1}{br}\sum_i \left[\sum_j \sum_k y_{ijk}\right]^2 - C = \frac{(88)^2 + (77.5)^2 + (95)^2}{12} - C.$$

$$= 1897.93 - 1885.00 = 12.94.$$

$$SCE = \frac{1}{r}\sum_i \sum_j \left[\sum_k y_{ijk}\right]^2 - \frac{1}{br}\sum_i \left[\sum_j \sum_k y_{ijk}\right]^2.$$

$$= \frac{1}{4}\left[(32)^2 + (28)^2 + \cdots + (35)^2\right] - 1\,897.93 = 1\,915.56 - 1\,897.93 = 17.63.$$

$$SCS = SCT - SCA - SCE = 67.75 - 12.94 - 17.63 = 37.18$$

La variación entre materas es no significativa ($p = 0.083$) y una estimación de su valor es:

$$\hat{\sigma}^2 = \frac{MCE - MCS}{r} = \frac{2.94 - 1.38}{4} = 0.39.$$

La variación total es la suma de las varianzas entre plantas y entre materas, esto es, $\hat{\sigma}_s^2 + \hat{\sigma}^2 = 1.38 + 0.39 = 1.77$. La variación entre materas es el 22% de la variación total de cada observación. Este resultado indica que en una investigación posterior deberá concentrarse la atención en plantas más que en materas.

Las diferencias con respecto a las horas de luz diurna tampoco son significativas ($p = 0.19$). Para esta prueba se usó como denominador la *MCE* que estima la suma $\sigma_s^2 + r\sigma^2$. Varios autores, entre ellos Gill (1978, vol. 1) y Sokal y Rohlf (1995), dicen que cuando la varianza σ^2 es no significativa, tanto la *MCS* como la *MCE* estiman la varianza σ_s^2, así que

Tabla 10.3 Tabla del ANOVA para los datos del ejemplo 10.1

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Horas de luz (tratamiento)	12.94	2	6.47	6.47 / 2.94 = 2.20	0.19
Entre materas (error experimental)	17.63	6	2.94	2.94 / 1.38 = 2.13	0.083
Entre plantas (error de muestreo)	37.18	27	1.38		
Total	67.75	35			

el promedio de ellas sería una mejor estimación de este error ya que estaría basada en un número mayor de *gl*.

Ahora bien, no existen criterios unánimes para decidir cuándo mancomunar las sumas de cuadrados, sin embargo Storm (1962) propuso una regla simple, aunque no muy rígida, que es la que se seguirá aquí.

Si $F = MCE \,/\, MCS < 2F_{0.50,\ \upsilon_1,\ \upsilon_2}$, entonces pueden promediarse las sumas de cuadrados como:

$$MCE' = \frac{SCS + SCE}{abr - a}.$$

En el ejemplo anterior, $F = 2.13$ es mayor que $2F_{0.50,\ 6,\ 24} = 1.83$; por tanto, no se promedian estas sumas. En los diseños anidados la *MCE* tiene asociados pocos *gl*, lo que es una desventaja, pues se reduce la potencia de las pruebas.

Una estimación de las medias de tratamientos es $\hat{\mu}_8 = 88/12 = 7.3$ cm $\hat{\mu}_{12} = 77.5/12 = 6.5$ cm, y $\hat{\mu}_{16} = 95/12 = 7.9$ cm; y una estimación de la media poblacional es $\hat{\mu} = 260.5/36 = 7.24$ cm.

La varianza de las medias de los tratamientos se estima con la fórmula:

$$Var(\bar{y}_i) = (1/br)MCE = 2.94/12 = 0.245.$$

Se advierte que la varianza anterior se calcula con base en la *MCE*, o sea la variación entre materas, y no con base en la variación entre plantas. Las fórmulas para intervalos de confianza y comparaciones entre medias tienen algunas variaciones, como lo indica Gill (1978, vol. 1).

No debe olvidarse el chequeo de los supuestos del ANOVA, principalmente el de homogeneidad de varianzas. Estas se calculan para los totales por matera y no para observaciones individuales. La fórmula es:

$$S_i^2 = [1 \,/\, rb((b-1)]\ [b\Sigma\, y_{ij.}{}^2 - (\Sigma\, y_{ij.})^2].$$

Luego:

$$S_1^2 = [1/4 \times 3 \times 2)\ \{3[(32)^2 + (28)^2 + (28^2)] - (88)^2\} = 1.333,$$

$$S_2^2 = 3.146 \text{ y } S_3^2 = 4.333.$$

Aplicando la prueba de Hartley, se tiene que $F_{máx} = 4.333 \,/\, 1.333 = 3.25 < 8.75 = f_{0.05,\ 3,\ 2}$. Los *gl* son $k = 3(= a)$ y $r - 1 = 2(= b - 1)$. Se acepta el supuesto de homoscedasticidad en los datos.

10.2 Modelo aleatorio

El diseño anidado o jerarquizado se caracteriza porque las clasificaciones subordinadas están anidadas dentro de un nivel de clasificación más alto. En el ejemplo

anterior, las materas están subordinadas a las horas de luz diurna. Ahora bien, los tratamientos de un factor subordinado deben elegirse aleatoriamente. Solamente el factor de la clasificación más alta puede ser de efectos fijos o de efectos aleatorios, y este factor es el que define el tipo de modelo.

En el modelo mixto, el submuestreo permite averiguar el tamaño del error en cada uno de los niveles de jerarquización del experimento. Por ejemplo, si se desea conocer el efecto de algunas drogas sobre la cantidad de pigmento en la piel de las ratas, la droga será el factor de clasificación más alto; el experimento es de efectos fijos si las drogas se escogen previamente. El experimento se realiza tomando al azar varias ratas para cada droga; de la región ventral de cada rata se toman al azar tres muestras de piel; por último, cada una de estas muestras se macera y se submuestrean porciones que se hidrolizan separadamente. La cantidad de pigmento se lee como una densidad óptica, lectura que puede repetirse varias veces en cada muestra hidrolizada.

En este ejemplo se tienen varios niveles de subordinación: ratas dentro de drogas, muestras de piel dentro de ratas, hidrolización dentro de muestras de piel y lecturas dentro de muestras hidrolizadas. La varianza del error se origina en las lecturas repetidas de la densidad óptica para cada muestra hidrolizada. El diseño anidado también permite estimar las variabilidades entre muestras hidrolizadas dentro de cada muestra de piel, entre muestras de piel de la misma rata y entre ratas asignadas a la misma droga. La magnitud de cada varianza es un indicador que muestra dónde debe aplicarse mayor control experimental y a qué niveles debe el experimentador enfocar los mayores esfuerzos.

El modelo II anidado o modelo aleatorio puro, tiene aplicación en varias ramas de la biología. Por ejemplo, en genética cuantitativa, su uso permite estimar los porcentajes de variación atribuibles a diferentes causas contempladas en el estudio y, a la vez, los coeficientes de correlación intraclase o de heredabilidad. En estudios de variación geográfica, es posible estimar las varianzas entre muestras de una localidad y entre localidades, para analizar patrones de distribución poblacionales de varios organismos.

El modelo matemático para un diseño anidado en dos niveles y efectos aleatorios tiene la siguiente forma:

$$y_{ijk} = \mu + A_i + B_{(i)j} + \delta_{(ij)k}, \qquad (10.2)$$

que es similar al modelo (10.1), pero el factor A es de efectos aleatorios, por lo que las A_i son variables aleatorias independientes con distribución $N(0,\sigma_A^2)$. La tabla del ANOVA para el modelo aleatorio es la misma tabla 10.1, cambiando la MC esperada del factor A por la suma $\sigma_s^2 + r\,\sigma^2 + br\,\sigma_A^2$.

Ejemplo 10.2. Se llevó a cabo un estudio piloto para caracterizar las fuentes de variación aleatoria en la rata de sedimentación de eritrocitos, en bebés de un día de nacidos. Para este estudio se eligieron al azar tres hospitales y de cada uno de ellos se tomó una muestra, también al azar, de tres sujetos. Las lecturas se hicieron por triplicado para cada bebé y los resultados se muestran en la tabla 10.4 (adaptado de Gill, 1978, vol. 1).

Las varianzas muestrales entre sujetos para cada hospital son 1811.11, 2036.11 y 811.11, respectivamente, y pueden considerarse homogéneas sin necesidad de aplicarles una prueba formal.

Tabla 10.4 Datos correspondientes al ejemplo 10.2

Hospitales	1			2			3		
Sujetos	1	2	3	4	5	6	7	8	9
Lecturas	110	130	50	130	45	120	100	130	70
	90	115	75	45	55	50	200	80	80
	120	105	85	50	65	150	90	70	40
Totales	320	350	210	225	165	320	390	280	190
		(880)			(710)			(860)	

Tabla 10.5 Tabla del ANOVA para el ejemplo 10.2

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Entre hospitales	1918	2	959.0	0.40	0.69
Entre sujetos dentro de hospitales	14384	6	2397.3	1.98	0.12
Entre lecturas intra sujetos	21783	18	1210.2		
Total	38085	26			

Tanto el factor A —hospitales— como el factor B —sujetos— son aleatorios. Como las fórmulas para calcular las sumas de cuadrados son las mismas que las del modelo mixto, se tienen los siguientes resultados, con los cuales se construye la tabla 10.5.

$C = (2450)^2 / 27 = 222315.$

$SCT = [(110)^2 + (90)^2 + \cdots + (40)^2] - C = 38085.$

$SCA = (1/9)[(880)^2 + (710)^2 + (860^2)] - C = 224233 - C = 1918.$

$SCE = (1/3)[(320)^2 + \cdots + (190)^2] - 224233 = 14834.$

$SCS = 38085 - 1918 - 14384 = 21783.$

La prueba F para variabilidad entre hospitales (H_0: $\sigma_A^2 = 0$) es no significativa. Igualmente, resultó no significativa la prueba de variabilidad entre sujetos (H_0: $\sigma^2 = 0$).

La varianza por observación o lectura es la suma de tres varianzas: $\sigma_s^2 + \sigma^2 + \sigma_A^2$. Cada una de estas puede estimarse a partir de las medias cuadráticas, así:

$\hat{\sigma}_s^2 = MCS = 1210.2.$

$\hat{\sigma}^2 = (1/r)(MCE - MCS) = (1/3)(2397.3 - 1210.2) = 395.7.$

$\hat{\sigma}_A^2 = (1/br)(MCA - MCE) = (1/9)(959.0 - 2397.3) = -159.8.$

La última estimación resultó negativa, pero aunque puede serlo desde el punto de vista matemático, no es una estimación aceptable para una varianza. Por otro lado, este resultado indica que el número de grados de libertad asociado con al menos una media cuadrática, es pequeño e insuficiente para estimar las componentes de varianza.

Se podría calcular un intervalo de confianza para esta varianza y observar si el intervalo obtenido contiene el cero. En este caso, se concluiría que no hay variación entre los hospitales. Como el objetivo es ilustrar el análisis de las componentes de varianza, se tendrá en cuenta la conclusión obtenida de la tabla del ANOVA y se supondrá que esta varianza es cero.

La varianza total será entonces $1210.2 + 395.7 + 0 = 1605.9$. El porcentaje de

variabilidad correspondiente a lecturas o repeticiones es 100 × 1210.2 / 1605.9% = 75.4%. El porcentaje restante (24.6%) es la variación atribuible a la diferencia entre sujetos, ya que la variabilidad entre hospitales es mínima y se ha igualado a cero. En estudios posteriores sobre este mismo fenómeno, deberá tenerse en cuenta que la mayor variación se dio entre las lecturas sobre el mismo sujeto. El control puede hacerse ya sea cambiando el método o el aparato de medición por uno más preciso, o haciendo mayor número de lecturas por sujeto y aumentando el número de sujetos.

Con respecto a la discusión anterior, la pregunta fundamental está relacionada con la o las etapas del muestreo en las que deberá hacerse más énfasis. Este problema está ampliamente discutido en Sokal y Rohlf (1995), bajo el tema de la distribución óptima de recursos en una investigación. Mendenhall (1968) también trata este tema y además el problema relacionado con los intervalos de confianza para componentes de varianza siguiendo los procedimientos de Satterthwaite y Welch.

Por otro lado, una estimación insesgada de la media poblacional μ es $\hat{\mu} = 2450 / 27 = 90.7$. Un intervalo de confianza para esta media tiene la forma $\bar{y} \pm t_{\alpha/2,\upsilon} S(\bar{y})$, donde el error estándar se estima de $S^2(\bar{y}) = MCA/abr = 959/27 = 35.52$, con $\upsilon = a - 1$ *gl*. Para el 95% de confiabilidad, el intervalo es 90.7 ± 4.303 × 5.96 = (65, 116).

10.3 Otros modelos

En algunas aplicaciones experimentales el diseño puede resultar desbalanceado, sea porque no todas las submuestras son de igual tamaño o porque se planteó un número diferente de repeticiones. El desbalanceo presenta problemas sobre todo en la estimación de parámetros; muchas de las estimaciones de componentes de varianza resultan sesgadas y las fórmulas conocidas para calcular los respectivos *gl* son, con frecuencia, aproximadas.

Los modelos 10.1 y 10.2 corresponden a diseños anidados en tres etapas, es decir, hay tres fuentes de variación. Pero el principio de jerarquización puede extenderse a cuatro o más fuentes de variación anidadas, dependiendo de la situación experimental. Por ejemplo, en el caso de la evaluación de algunas drogas en el pigmento de la piel de las ratas, el experimento puede diseñarse en cinco etapas. Ejemplos y procedimientos numéricos para estos modelos se encuentran en libros de diseño experimental más extensos como los de Gill, Ostle y Sokal y Rohlf, ya citados.

Los experimentos anidados exigen diseños diferentes a los de los experimentos factoriales; sin embargo, pueden presentarse experimentos mixtos, parcialmente jerarquizados o de clasificación factorial anidada.

Si en un experimento con tres factores al menos uno de ellos es aleatorio, este puede estar anidado dentro de un segundo factor y, a la vez, cruzado con un tercero. Considérese, por ejemplo, la raza como un factor *A* fijo y las camadas como un factor *B* anidado en *A*. Un tercer factor *C* aleatorio podría ser un grupo de tratamientos. Entonces, el modelo matemático que permite estudiar este diseño es de la forma:

$$y_{ijkl} = \mu + \alpha_i + B_{(i)j} + C_k + (\alpha C)_{ik} + (BC)_{(i)jk} + \varepsilon_{(ijk)l}, \tag{10.3}$$

donde $B_{(i)j}$ representa el efecto del *j*-ésimo individuo dentro de la camada correspondiente a la *i*-ésima raza, α_i es el efecto del nivel *i* del factor *A*, C_k es el efecto aleatorio del nivel *k* del factor C, y $(\alpha C)_{ik}$ y $(BC)_{(i)jk}$ son las interacciones entre los

factores A y C y B y C, respectivamente. Nótese que no hay interacción entre A y B, puesto que B está anidado en A; tampoco existe la interacción triple ABC.

También es común encontrar investigaciones que originan experimentos factoriales con submuestreo. Supóngase que se asignan al azar c individuos a cada una de las combinaciones de a niveles de un factor A y b niveles de otro factor B. De cada individuo se toman r submuestras que constituyen las observaciones repetidas del experimento. El modelo matemático apropiado para este experimento, considerando los factores A y B fijos, es:

$$y_{ijkl} = \mu + \alpha_i + B_j + (\alpha\beta)_{ij} + C_{(ij)k} + \varepsilon_{(ijk)l}. \qquad (10.4)$$

La tabla del ANOVA para analizar este modelo tiene la estructura de la tabla 10.6.

Ejemplo 10.3 Para investigar la concentración (en mg/100 mL) de calcio en el plasma de una especie de aves, se tomaron como factores el sexo y un tratamiento hormonal. A cada combinación de tratamientos se le asignaron tres aves y de cada una de ellas se tomaron dos muestras de sangre; así que las aves individuales son las muestras —factor anidado— y las lecturas, por duplicado, en la sangre son las submuestras (adaptado de Zar, 1999).

El factor A —sexo— tiene dos niveles: a_0 para hembras y a_1 para machos. El factor B —tratamiento hormonal— tiene niveles b_0 para no aplicado y b_1 para aplicado. Las aves conforman el factor C, el cual está anidado en las combinaciones de los factores A y B. Los datos figuran en la tabla 10.7.

Primero se calcula la varianza de cada uno de los tratamientos para probar homogeneidad.

$$s_1^2 = (1/12)\{3[(34.9)^2 + (26.7)^2 + (27.8)^2] - (89.4)^2\} = 9.905,$$

$$s_2^2 = 0.902, s_3^2 = 45.905, s_4^2 = 9.875.$$

La prueba de Burr-Foster permite concluir homogeneidad, puesto que $q = 0.52 < 0.72 = q_{0.01,2,4}$.

Las sumas de cuadrados de la tabla del ANOVA, tabla 10.9, se obtienen de los datos de la tabla 10.7 para el factor anidado y de la tabla 10.8, para los factores cruzados.

Tabla 10.6 Tabla del ANOVA para analizar el modelo (10.4)

Fuente de variación	*SC*	*gl*	*MC*	*MC esperada*	*F*
Factor A	*SCA*	$a-1$	*MCA*	$\sigma_s^2 + r\sigma^2 + \frac{bcr}{a-1}\Sigma\alpha_i^2$	$\frac{MCA}{MCE}$
Factor B	*SCB*	$b-1$	*MCB*	$\sigma_s^2 + r\sigma^2 + \frac{acr}{b-1}\Sigma\beta_j^2$	$\frac{MCB}{MCE}$
Interacción AB	*SCAB*	$(a-1)(b-1)$	*MCAB*	$\sigma_s^2 + r\sigma^2 + \frac{cr}{(a-1)(b-1)}\Sigma(\alpha\beta)_{ij}^2$	$\frac{MCAB}{MCE}$
Factor C (entre subgrupos)	*SCE*	$ab(c-1)$	*MCE*	$\sigma_s^2 + r\sigma^2$	$\frac{MCE}{MCS}$
Dentro de subgrupos	*SCS*	$abc(r-1)$	*MCS*	σ_s^2	
Total	*SCT*	$abcr-1$			

$C = (533.5)^2 / 24 = 11\,859.26.$

$SCT = (16.5)^2 + \cdots + (33.0)^2 - C = 14\,133.31 - 11\,859.26 = 2\,274.05.$

$SCB = (1 / 12)[(163.0)^2 + (370.5)^2] - C = 1\,794.01.$

$SCA = (1 / 12)[(288.0)^2 + (245.5)^2] - C = 75.26.$

$SCAB = (1 / 6)[(89.4)^2 + (73.6)^2 + (198.6)^2 + (171.9)^2] - C - SCA - SCB = 4.95.$

$SCE = (1 / 2)[(34.9)^2 + \cdots + (62.3)^2] - (1 / 6)[(89.4)^2 + \cdots + (171.9)^2]$
$= 13\,866.66 - 13\,733.48 = 133.18.$

$SCS = SCT - (SCA + SCB + SCAB + SCE) = 266.65.$

Los resultados de la tabla anterior permiten llegar a las siguientes conclusiones:

1. No hay interacción entre el sexo y el tratamiento hormonal.

2. No hay diferencias entre sexos con respecto a la concentración del calcio en la sangre.

3. El tratamiento hormonal es efectivo, pues aumenta significativamente la concentración de calcio en la sangre de las aves. La concentración promedio sin tratamiento hormonal es: $\hat{\mu}_0 = 163/12 = 13.6$ mg/100 mL y con el tratamiento es $\hat{\mu}_1 = 370.5/12 = 30.9$ mg/100 mL.

Tabla 10.7 Tabla de datos correspondientes al ejemplo 10.3

Tratamientos	*Aves*	*Repeticiones*	*Totales*
	1	16.5, 18.4	34.9
a_0b_0	2	12.7, 14.0	26.7
	3	12.8, 15.0	27.8
	4	14.5, 11.0	25.5
a_1b_0	5	10.8, 14.3	25.1
	6	10.0, 13.0	23.0
	7	39.1, 26.2	65.3
a_0b_1	8	21.3, 35.8	57.1
	9	40.2, 36.0	76.2
	10	32.0, 23.8	55.8
a_1b_1	11	28.8, 25.0	53.8
	12	29.3, 33.0	62.3

Tabla 10.8 Datos de la tabla 10.7 para los factores A y B cruzados

Factor A	*Factor B*		*Totales*
	b_0	b_1	
a_0	34.9	65.3	288.0
	26.7 (89.4)	57.1 (198.6)	
	27.8	76.2	
a_1	25.5	55.8	245.5
	25.1 (73.6)	53.8 (171.9)	
	23.0	62.3	
Totales	163.0	370.5	533.5

Tabla 10.9 Tabla del ANOVA para el ejemplo 10.3

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Sexo (A)	75.26	1	75.26	4.52	0.066
Tratamiento hormonal (B)	1794.01	1	1794.01	107.75	<0.001
Sexo por tratamiento hormonal (AB)	4.95	1	4.95	0.30	0.60
Aves (en AB)	133.18	8	16.65	0.75	0.65
Repeticiones en aves	266.65	12	22.22		
Total	2274.05	23			

4. La variabilidad entre submuestras o lecturas de la misma ave se estima mediante $\hat{\sigma}_s^2 = MCS = 22.22$. La variabilidad entre aves se estima como $\hat{\sigma}^2 = (1/r)(MCE - MCS)$. Esta estimación resulta negativa, entonces una alternativa es aceptar que la variabilidad entre aves puede ser cero. Los diseños anidados que se han estudiado son del tipo completamente aleatorio. Pueden diseñarse, igualmente, experimentos anidados y en bloques aleatorizados; su análisis sería similar al de los modelos 10.1 y 10.2, teniendo en cuenta la fuente de variación adicional atribuible a los bloques. Ejemplos de estos modelos se encuentran en Ostle (1988).

Ejercicios

10.1 Se planeó un diseño anidado con la finalidad de evaluar la variabilidad en la concentración de ácido ascórbico (mg/100 mL) en plantas de nabos. Inicialmente se seleccionaron cinco plantas de un cultivo —una parcela— en forma aleatoria y de la parte central de la planta se muestrearon dos hojas. Luego se determinó, por triplicado, la concentración de ácido ascórbico en cada hoja y se obtuvieron los resultados que siguen (adaptado de Ostle, 1988).

Plantas	1		2		3		4		5	
Hojas	1	2	3	4	5	6	7	8	9	10
Repeticiones	9.1	7.3	7.3	9.0	7.3	8.9	10.7	12.7	7.7	9.4
	7.3	9.0	6.6	8.4	5.2	6.9	5.3	6.8	6.7	8.3
	10.8	12.5	9.3	11.0	7.3	8.9	9.3	11.2	10.4	12.0

a. Establecer el modelo y los supuestos para analizar este diseño.
b. Construir la tabla del ANOVA respectivo.
c. Enunciar las hipótesis estadísticas y realizar el análisis de significación para cada prueba.
d. Estimar los porcentajes de variación atribuibles a diferencias entre plantas, entre hojas y entre repeticiones.
e. Discutir la elección de los tamaños muestrales para un estudio similar que se vaya a realizar.
f. Calcular un intervalo de confianza del 95% para la media general μ.

10.2 Para analizar el pH en la sangre de ratas, estas se seleccionaron de las camadas producidas por el apareamiento de tres padres diferentes con

cada una de cuatro madres. Las lecturas correspondientes a cuatro ratas por camada fueron las siguientes (Sokal y Rohlf, 1995).

Madres	*Padres*	*Repeticiones* 1	2	3	4
1	1	7.40	7.45	7.42	7.48
	2	7.45	7.33	7.40	7.46
	3	7.40	7.47	7.40	7.47
2	4	7.44	7.51	7.49	7.51
	5	7.49	7.49	7.49	7.50
	6	7.48	7.59	7.59	7.52
3	7	7.39	7.37	7.33	7.43
	8	7.43	7.38	7.44	7.42
	9	7.46	7.44	7.47	7.54
4	10	7.47	7.49	7.45	7.43
	11	7.45	7.42	7.52	7.51
	12	7.51	7.51	7.53	7.45

a. Construir y analizar la tabla del ANOVA para estos datos.
b. Estimar las varianzas entre madres, entre padres y entre ratas de la misma camada.
c. Dar sugerencias sobre tamaños muestrales con base en los resultados obtenidos en *b.*

10.3 En una investigación sobre el contenido urinario de potasio en hombres y mujeres sanos, se seleccionó al azar una muestra de nueve personas de cada sexo; de cada uno de ellos se efectuaron tres lecturas en condiciones basales (adaptado de material mimeografiado, curso de bioestadística, Universidad de Chile, 1971. Sin autor).

Hombres				*Mujeres*			
Persona	*Lecturas*			*Persona*	*Lecturas*		
1	48	45	42	1	69	59	63
2	29	31	35	2	22	21	21
3	50	53	50	3	94	86	76
4	67	59	70	4	58	51	43
5	110	102	75	5	46	46	40
6	56	57	52	6	65	62	63
7	91	121	98	7	57	54	47
8	53	58	51	8	96	78	69
9	114	96	96	9	138	128	125

a. Establecer el modelo matemático y sus supuestos.
b. Construir la tabla del ANOVA y probar las hipótesis del problema.
c. Estimar la media del contenido urinario de potasio para cada sexo. ¿Tiene sentido estimar una media general?
d. Estimar el porcentaje de variabilidad entre personas.
e. Realizar una prueba de homogeneidad de varianzas para el ANOVA.

10.4 Supóngase una floresta en la que se ha obstruido parcialmente su flujo hídrico a causa de la construcción de una carretera. Transcurrido un tiempo suficiente para que la especie más representativa exhiba cambios poblacionales, se toman muestras de su densidad sobre cuatro estaciones seleccionadas aleatoriamente a cada lado de la vía. Los valores correspondientes a dichas muestras se precisan en la siguiente tabla (tomado de Ramírez, 1988).

Zonas	*1*				*2*			
Estaciones	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Muestreos	12	14	9	8	8	2	5	5
	10	7	11	10	8	3	6	2
	18	12	10	10	11	0	5	3
	16	9	10	13	7	4	4	2
	13	7	10	15	9	2	6	2
	20	12	6	11	8	3	6	3

a. Construir la tabla del ANOVA para analizar estos datos.
b. Comparar zonas y estimar sus valores promedio.
c. Hallar la variabilidad entre estaciones y entre muestras.
d. Discutir la siguiente conclusión: un estudio futuro similar deberá contemplar menos muestras por estación y, a su vez, un mayor número de estaciones.

10.5. Para estudiar el efecto de tres niveles de digitalina, factor *A*, sobre el nivel de calcio en el músculo del corazón, se eligieron perros de dos razas, factor *B*; de cada músculo se tomaron dos muestras al azar. Las lecturas de cada muestra se hicieron por duplicado, como se muestra en la siguiente tabla (adaptado de Mendenhall, 1968).

Tratamientos	*Muestras de tejidos*	*Repeticiones*	
a_0b_0	1	1342	1335
	2	1608	1603
a_1b_0	3	1881	1890
	4	1698	1702
a_2b_0	5	1387	1384
	6	1140	1148
a_0b_1	7	1296	1298
	8	1029	1034
a_1b_1	9	1549	1546
	10	1150	1151
a_2b_1	11	1579	1571
	12	1319	1324

a. Establecer el modelo matemático y los supuestos para analizar estos datos.
b. Construir la tabla del ANOVA y obtener las conclusiones sobre los factores.
c. Estimar la variabilidad entre muestras y entre repeticiones.

11

Regresión lineal simple

11.1 Concepto de regresión

El modelo matemático de un diseño experimental relaciona una variable de respuesta generalmente cuantitativa, llamada variable dependiente, con uno o más factores de estudio, o sea las variables independientes. Los diseños analizados en los capítulos anteriores suponen que los factores son variables cualitativas que pueden descomponerse en un número finito de niveles o tratamientos. Si el factor es cuantitativo, sus niveles deben escogerse en puntos fijos de la escala o rango de valores. Por ejemplo, cuando la temperatura entra como un factor de estudio, podría tomarse como niveles los valores 20, 25 y 30 ºC. El ANOVA compara estos tres niveles pero no tiene en cuenta los valores intermedios entre los intervalos 20-25 y 25-30 ºC.

La metodología estadística llamada análisis de regresión fue desarrollada inicialmente para analizar la relación estadística entre una variable dependiente y una o más variables independientes; todas ellas cuantitativas y definidas en un rango de variación adecuado. Actualmente, el análisis de regresión se aplica también a las variables discretas y cualitativas, aunque con enfoques diferentes a los tradicionales.

Cuando el investigador sospecha la existencia de alguna relación o asociación entre variables medibles, posiblemente del tipo causa-efecto, la mejor manera de estudiarla es estableciendo una relación funcional de tipo matemático entre ellas. La variable de respuesta Y se llamará variable dependiente y las variables explicativas de Y serán las variables predictoras o independientes, denotadas por $X_1, X_2 \dots X_p$.

Uno de los objetivos del análisis de regresión es hallar una función que relacione a Y con sus variables explicativas y que pueda escribirse como $Y = f(X_1, X_2 \dots X_p)$. Por ejemplo, la producción de un árbol frutal depende de varios factores: la calidad de la semilla, la fertilidad del suelo y la influencia del clima, entre otros; la estatura de una persona depende de factores genéticos, como las estaturas de sus padres, y de factores ambientales como la alimentación, el ejercicio físico, etc. En estos ejemplos, tanto la producción como la estatura pueden estudiarse como funciones estadísticas de las otras variables o factores.

La forma de la función f es desconocida para el investigador, pero en muchos casos puede suponerse una función polinómica de la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p. \quad (11.1)$$

Cuando se considera una sola variable explicativa, la función tiene la forma:

$$Y = \beta_0 + \beta_1 X. \quad (11.2)$$

Las ecuaciones (11.1) y (11.2) representan modelos determinísticos para la variable dependiente. Si se desea analizar una relación estadística entre las variables, es necesario agregar una componente aleatoria ε para denotar el error de medición en la variable de respuesta. Entonces, el modelo estadístico puede escribirse como;

$$Y = \beta_0 + \beta_1 X + \varepsilon. \quad (11.3)$$

Esta ecuación representa el modelo de regresión lineal simple de primer orden entre la variable aleatoria Y y una variable no aleatoria X. La linealidad se refiere a los coeficientes beta llamados parámetros del modelo, los cuales deben tener la unidad como exponente y figurar como coeficientes individuales, en otras palabras, la variable Y se expresa como una combinación lineal de los parámetros β_i. El modelo es simple cuando contiene una sola variable explicativa y múltiple cuando se incluyen dos o más variables explicativas. Se dice que es un modelo de primer orden porque la variable X tiene exponente igual a uno. El modelo de regresión lineal simple de segundo orden, tiene la forma:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon.$$

Un modelo de regresión es no lineal cuando no hay linealidad en los parámetros, por ejemplo $Y = \beta_0 + X^{\beta_1} + \varepsilon$ o $Y = 1/\left[\beta_0 + e^{\beta_1 X}\right] + \varepsilon$, aquí el parámetro β_1 figura como exponente y no como coeficiente de un sumando.

Por razones de extensión y complejidad, aquí sólo se estudiará el modelo de regresión lineal simple de primer orden. Para el estudio y aplicación de otros modelos, véanse Zar (1999) y Quinn y Keough (2002). Modelos más avanzados en la investigación biomédica se encuentran en Harrel (2001) y Dupont (2002).

En estudios experimentales, la variable X es no aleatoria: el investigador puede manipularla, o sea, elegir los valores que va a tomar en el experimento de acuerdo con la conveniencia y facilidad en su manejo. En estudios observacionales, la unidad experimental ofrece a la vez los valores de X y de Y, de modo que la variable X también debe considerarse aleatoria.

El término regresión se debe a Francis Galton, quien lo usó por primera vez en una publicación en 1885 sobre las leyes de la herencia ("Regression towards mediocrity in hereditary stature"). En sus investigaciones encontró que algunas características de los hijos se asemejaban a las de los padres pero *regresaban*, en cierta medida, a los estándares de la raza. Así, si los padres son altos, los hijos también tienden a serlo, aunque su estatura está más ajustada al promedio de la raza, o sea, *regresa* o retrocede al estándar. Igualmente, los hijos de padres de baja estatura tienden a ser bajos, pero

su estatura de nuevo *regresa* al promedio racial. El comportamiento específico de la estatura fue verificado por K. Pearson en 1903 para Inglaterra.

La palabra regresión perdió su significado original y con el tiempo adquirió el sentido de predicción o relación estadística entre variables, con el cual se conoce actualmente.

11.2 Regresión lineal simple

Cuando se estudia el modelo lineal simple, se hace referencia al modelo (11.2). Esta ecuación puede representarse gráficamente mediante una línea recta en la cual a cada valor de X le corresponde un solo valor de Y. No obstante, en la práctica la variable Y es aleatoria, de modo que para un valor particular de X generalmente se obtienen diferentes valores de la respuesta Y. Entonces, para establecer la relación lineal entre las variables X y Y, se supone que a cada valor de X le corresponde el valor promedio o valor esperado de Y, escrito $E(Y / X = x) = \mu_{y.x}$. Así, la regresión lineal tiene la forma:

$$\mu_{y.x} = \alpha + \beta X. \tag{11.4}$$

Esta expresión es el modelo poblacional que representa la relación lineal entre las dos variables, por tanto, los parámetros $\mu_{y.x}$, α y β son cantidades desconocidas que es necesario estimar en el modelo. En adelante, se usará α en vez de β_0 y β en vez de β_1 con el fin de simplificar la escritura.

En las investigaciones diseñadas para analizar regresión entre dos variables, debe obtenerse una muestra de n pares de valores (x_i, y_i) a partir de la cual podrán estimarse los parámetros de la regresión. Los valores observados y_i son diferentes valores de la variable aleatoria Y, los cuales se desvían de su correspondiente valor promedio en una cantidad ε_i, llamada error aleatorio (véase figura 11.1).

En términos de los valores observados, el modelo de regresión se representa de esta forma:

$$y_i = \alpha + \beta x_i + \varepsilon_i; \text{ con } i = 1, 2, \ldots, n. \tag{11.5}$$

Las inferencias en regresión simple se basan en cuatro supuestos sobre los elementos de la ecuación (11.5). Estos son:

1. La variable explicativa X se mide sin error, es decir, está bajo el control del investigador quien escoge los valores sin las influencias del azar.

2. La relación entre los valores esperados de Y y los valores de X puede expresarse mediante una función lineal.

3. Los diferentes valores de Y para cada valor de X son independientes y tienen distribución Normal. Otra forma de describir este supuesto es exigir que los errores ε_i sean probabilísticamente independientes y con distribución $N(0,\sigma^2)$ para cada valor de X.

4. Los errores ε_i tienen varianzas homogéneas para cada valor de X, o sea que la varianza en torno a la línea de regresión es constante y no depende del valor que tome X.

La metodología estadística parte de un modelo que representa la situación poblacional, esto es, el comportamiento de las variables en el fenómeno natural. Luego selecciona una muestra aleatoria del fenómeno y, con estos datos, construye una representación o imagen del modelo teórico. Finalmente, analiza el modelo muestral y lo acepta o rechaza dependiendo de la calidad del ajuste obtenido.

Con los pares de valores (x_i, y_i) se obtiene el modelo muestral correspondiente al modelo (11.4); dicho modelo se llama ecuación de regresión ajustada y su gráfica es una línea recta similar a la poblacional. Para obtener la ecuación ajustada se requiere una estimación de los coeficientes α y β; el procedimiento matemático usual para obtenerla es el método de los mínimos cuadrados, consistente en minimizar la suma de errores cuadráticos en la muestra aleatoria.

Para cada x_i, el error ε_i se define como la diferencia entre el valor observado y su correspondiente valor esperado, esto es, $\varepsilon_i = y_i - \mu_{y.x}$. Entonces, las mejores estimaciones de α y β son aquellos números a y b que hacen mínima la suma siguiente:

$$L = \sum_i \varepsilon_i^2 = \sum_i (y_i - \alpha - \beta x_i)^2.$$

Los cálculos aritméticos se realizan partiendo de la muestra aleatoria y obteniendo primero las siguientes sumas: $\sum x_i$, $\sum x_i^2$, $\sum y_i$, $\sum y_i^2$, $\sum x_i y_i$. Luego se calculan las sumas cuadráticas con las fórmulas siguientes:

$$S_{xx} = \sum x_i^2 - (1/n)(\sum x_i)^2.$$

$$S_{yy} = \sum y_i^2 - (1/n)(\sum y_i)^2.$$

$$S_{xy} = \sum x_i y_i - (1/n)(\sum x_i)(\sum y_i).$$

Los mejores estimadores de α y de β son $\hat{\alpha} = a$ y $\hat{\beta} = b$, que se calculan como:

$$b = S_{xy} / S_{xx}.$$

$$a = (1/n)\left(\sum y_i - b\sum x_i\right) = \bar{y} - b\bar{x}. \qquad (11.6)$$

Entonces, la ecuación de regresión muestral, llamada ecuación ajustada, es,

$$\hat{y} = a + bx. \qquad (11.7)$$

Para cada valor x_i se tiene un valor estimado, ajustado o pronosticado de Y, dado por $\hat{y}_i = a + bx_i$.

Un valor observado y_i, se representa en esta ecuación ajustada como:

$$y_i = a + bx_i + e_i. \qquad (11.8)$$

Los e_i *se* llaman residuos y son las diferencias entre los valores observados y los ajustados, esto es, $e_i = y_i - \hat{y}_i$. Cada e_i es un estimador del correspondiente error ε_i. La recta de regresión ajustada pasa por el punto $(\bar{x}, \bar{y})$, de modo que la ecuación ajustada también puede escribirse como:

$$\hat{y}_i = \bar{y} + b(x_i - \bar{x}). \qquad (11.9)$$

De otro lado, el modelo de regresión lineal simple es una ecuación de primer grado en X que puede representarse gráficamente con una línea recta en el plano cartesiano. En la figura 11.1 se dibuja una recta poblacional desconocida y una estimación de dicha recta basada en una muestra.

Si el modelo postulado es correcto, es decir, si la recta poblacional representa la verdadera relación entre las variables, la recta ajustada deberá ser similar a la poblacional o teórica y $e_i \approx \varepsilon_i$ para todo i. Ahora bien, para determinar si la recta estimada tiene buena precisión, esto es, si está cerca de la recta verdadera en la población, la estadística supone que los datos para la regresión se obtuvieron aleatoriamente, sin sesgos y sin errores sistemáticos. Se trata, entonces, de estimar la precisión de la recta ajustada mediante una medida de variabilidad; la más indicada es la varianza de los residuos e_i y mientras más pequeña sea esta varianza mejor es el ajuste. Gráficamente, el ajuste es mejor mientras más cerca estén los puntos observados de la recta de regresión.

De otro lado, la calidad del ajuste de una recta de regresión se mide por la varianza residual σ^2 y esta se estima mediante la *MCE* del ANOVA de regresión. Como el modelo de regresión es también un modelo lineal, se puede usar la técnica del ANOVA para estimar la varianza y probar la hipótesis de regresión.

La variabilidad en Y dada por $SCT = \sum (y_i - \bar{y})^2$ puede subdividirse en dos componentes: una, explicada por la relación de Y con X, llamada suma de cuadrados debida a la regresión, $SCR = \sum (\hat{y}_i - \bar{y})^2$; y la suma de cuadrados del error aleatorio, llamada suma de cuadrados alrededor de la regresión $SCE = \sum (y_i - \hat{y}_i)^2$. Esta última suma se origina porque no todos los puntos observados caen sobre la recta de regresión. Cada suma de cuadrados tiene asociado su propio número de gl, como se muestra en la tabla 11.1 correspondiente al ANOVA para la regresión lineal simple.

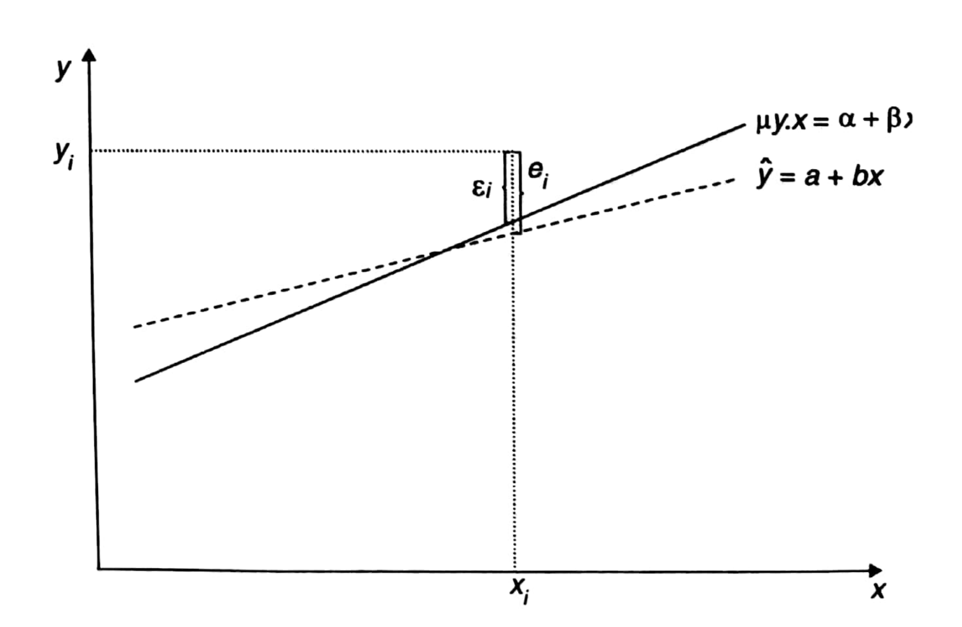


Figura 11.1 Error ε_i en el modelo poblacional (—) y residuos e_i en el modelo muestral (- - -) para un valor y_i.

La hipótesis básica de la regresión se refiere a la existencia de una relación lineal entre Y y X. Se dice que existe regresión de Y sobre X si se puede relacionar Y con X mediante la ecuación $Y = \alpha + \beta X$. En otras palabras, Y está relacionada linealmente con X si el coeficiente de regresión β es diferente de cero. Las hipótesis estadísticas se plantean de la siguiente manera:

H_0: $\beta = 0$ contra H_1: $\beta \neq 0$.

La regla de decisión consiste en rechazar H_0 si $F \geq F_{\alpha,1,n-2}$. En este caso se concluye que el término βX hace parte del modelo de regresión propuesto y, por tanto, la ecuación $y = a + bx$ es un buen modelo para explicar la relación lineal.

En las investigaciones biológicas no es usual obtener repeticiones de Y por cada valor X. Aunque este es un requerimiento teórico, es posible obtener la ecuación ajustada, estimar parámetros y hacer un diagnóstico de la regresión sin tener repeticiones de Y. Los ejemplos siguientes ilustran estos procedimientos y muestran resultados siguiendo los patrones que se encuentran frecuentemente en la literatura científica.

Ejemplo 11.1 Se requiere conocer la relación entre la concentración de una sustancia y su densidad óptica. La variable independiente X es la concentración, la cual se tomó a diferentes niveles desde 80 hasta 520 ppm. La variable de respuesta Y representa unidades de densidad óptica, DO, medidas para cada una de las concentraciones. Se tienen los siguientes datos:

Concentración X:	80	120	160	200	240	280	320	360	400	440	480	520
DO Y:	0.08	0.12	0.18	0.21	0.28	0.28	0.38	0.40	0.42	0.50	0.52	0.60

El primer paso en el análisis de regresión es graficar los puntos correspondientes a la muestra en los ejes cartesianos, para obtener una *nube de puntos o diagrama de dispersión* que guíe en el tipo de modelo que deberá postularse —lineal, cuadrático, exponencial, etc.

La figura 11.2 muestra el diagrama de puntos con una clara tendencia lineal, lo que sugiere la idea de ajustar los datos a una línea recta. Los coeficientes de regresión α y β se estiman con la fórmulas (11.6), así:

$$S_{xx} = 1308800 - (3600)^2 / 12 = 228800.$$
$$S_{yy} = 1.6153 - (3.97)^2 / 12 = 0.3019.$$
$$S_{xy} = 1452.4 - (3600)(3.97) / 12 = 261.4.$$

Entonces $b = 261.4 / 228800 = 0.00114$.

$$a = (3.97 - 0.00114 \times 3600) / 12 = -0.01191,$$

Tabla 11.1 Tabla del ANOVA para la regresión lineal simple

Fuente de variación	*SC*	*gl*	*MC*	*MC esperada*	*F*
Debida a la regresión (modelo)	*SCR*	1	MCR	$\sigma^2 + \beta^2 S_{xx}$	$\frac{MCR}{MCE}$
Alrededor de la regresión (error aleatorio)	*SCE*	$n-2$	MCE	σ^2	
Total	*SCT*	$n-1$			

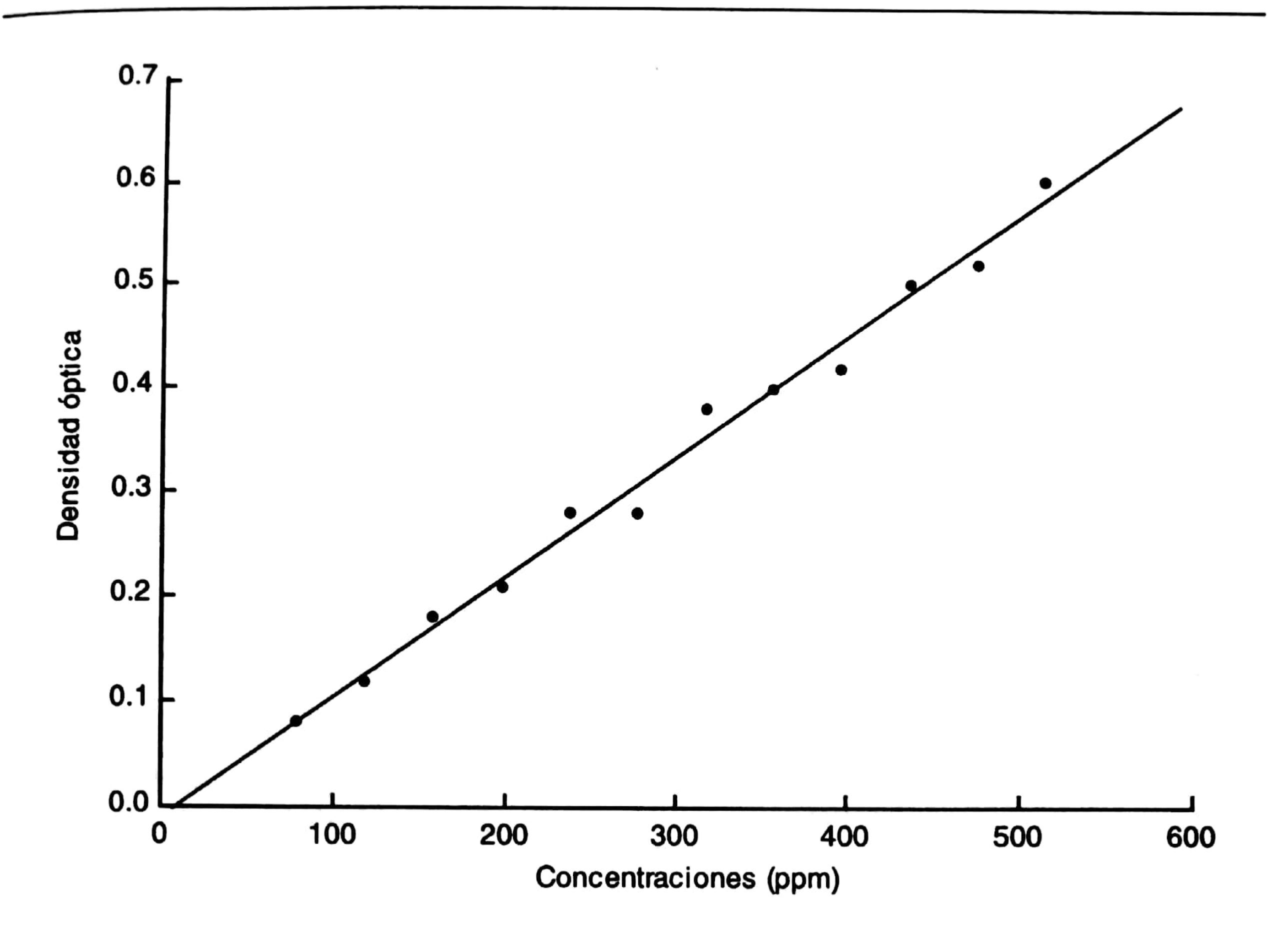


Figura 11.2 Diagrama de puntos para el ejemplo 11.1

y la recta ajustada es:

$$\hat{y}_i = -0.01191 + 0.00114x_i. \qquad (11.10)$$

Las sumas de cuadrados para la tabla del ANOVA (véase tabla 11.2) se calculan con las fórmulas:

$SCT = S_{yy} = 0.3019.$

$SCR = S_{xy}{}^2 / S_{xx} = (261.4)^2 / 228800 = 0.2986.$

$SCR = bS_{xy} = b^2 S_{xx}$; también puede usarse cualquiera de estas fórmulas.

$SCE = SCT - SCR = 0.3019 - 0.2986 = 0.0033.$

La prueba F del ANOVA es altamente significativa, se concluye, entonces, que Y está relacionada linealmente con X.

La variación de la DO está asociada a los niveles de concentración de la sustancia. Esta asociación o dependencia se mide con el *coeficiente de determinación* R^2 definido como:

$$R^2 = \frac{SCT - SCE}{SCT}.$$

En general, este coeficiente mide la reducción proporcional de la varianza total en Y al usar el estimador $\hat{y}$ (o sea el modelo completo, incluyendo el término βX) en vez de $\overline{y}$ (o sea el modelo reducido, haciendo $\beta = 0$) para estimar valores de Y.

Un valor alto de R^2 no indica que la relación de Y con X sea necesariamente lineal,

Tabla 11.2 Tabla del ANOVA para la regresión del ejemplo 11.1

Fuente de variación	*SC*	*gl*	*MC*	*F*	*P*
Debido a la regresión	0.2986	1	0.2986	904.8	<0.001
Error	0.0033	10	0.00033		
Total	0.3019	11			

es decir, no implica que X sea un buen predictor lineal de Y, solo que hay mayor precisión en la estimación de Y si se tiene en cuenta la variable X que si no se tiene en cuenta, Anderson-Sprecher (1994).

El coeficiente R^2 aumenta con el espaciamiento de los valores de X aunque el ajuste lineal no sea bueno. Por otra parte, $R^2 \approx 0$ puede indicar que no hay dependencia lineal de Y con X o que la dependencia es no lineal. Así, R^2 no mide idoneidad del modelo lineal.

Para el ejemplo anterior: $R^2 = (0.3019 - 0.0033) / 0.3019 = 0.9891$, expresado en porcentaje, permite afirmar que el 98.9% de la variación en la DO es explicada por su relación con el nivel de concentración de la sustancia. Este porcentaje de varianza explicada es alto y sugiere que es mejor usar el modelo lineal para estimar valores de Y que no usarlo. El 1.1% es varianza no explicada por la variable X sino por otras variables o causas que no se tuvieron en cuenta en el análisis.

Volviendo al ejemplo, se puede concluir que la ecuación de regresión ajustada pronostica, con menos error, la DO correspondiente a una concentración dada, que el modelo reducido, en este caso la media aritmética $\overline{y}$.

Suponiendo una concentración de 400 ppm, el modelo completo pronostica la DO como $\hat{y}$ de $= -0.01191 + 0.00114 \times 400 = 0.44$, mientras que el pronóstico dado por la media aritmética es 0.33, un valor más lejano de 0.42, que es el valor real.

11.3 Estimación de los parámetros de la regresión

En el análisis de regresión lineal simple se tienen como parámetros fundamentales los coeficientes α y β, la varianza σ^2 del error y el valor medio de Y para un X dado $\mu_{y.x}$. Sus estimaciones puntuales, considerando el ejemplo anterior, son:

1. $\hat{\alpha} = a = -0.01191$ representa el intercepto de la recta sobre el eje vertical. Tiene interpretación práctica solo cuando el cero está en el rango de X.

2. $\hat{\beta} = b = 0.00114$ representa la pendiente de la recta, o sea la variación en Y por unidad de cambio en X. En el ejemplo, por cada unidad de aumento en la concentración de la sustancia, la correspondiente DO aumenta 0.00114 unidades. Si b fuera negativa, entonces, al aumentar X en una unidad, la variable Y disminuiría en b unidades. La pendiente b tiene las unidades de Y / X.

3. $\hat{\sigma}^2 = MCE = 0.00033$. Una varianza mayor que cero quiere decir que en la práctica no existe una relación perfecta entre Y y X: se obtienen valores de Y que fluctúan alrededor de una recta. Esta fluctuación está descrita por los errores ε_i cuya varianza se estima con la MCE. Se supone además que σ^2 es constante para todos los valores de X. A veces la MCE se representa también por s^2.

4. $\hat{\mu}_{y.x} = \hat{y}$. La igualdad implica que la ecuación de regresión puede usarse como una ecuación de predicción de valores medios de Y. Para un valor de la concentración X dentro de su rango de validez, se puede estimar el valor promedio de la DO mediante $\hat{y}$. Así, para $X = 310$, la DO promedio será $\hat{y} = 0.34$. El pronóstico es válido

para valores de X en el rango elegido u obtenido en el muestreo. La extrapolación, es decir, tomar valores por fuera del rango de X, puede llevar a estimaciones erróneas. Por ejemplo, si la concentración es cero, la DO promedio es negativa, resultado que no es interpretable en la práctica.

También pueden estimarse los parámetros de la regresión usando intervalos confidenciales. Por ejemplo, un intervalo de confianza para β tiene la forma:

$$b \pm t_{\alpha/2,\, n-2} S_b. \tag{11.11}$$

El error estándar de b se obtiene de $S_b^2 = MCE / S_{xx} = 0.00033 / 228\,800 = 14.4 \times 10^{-10}$, luego $S_b = 3.8 \times 10^{-5}$. Entonces, un intervalo del 95% de confianza será $0.00114 \pm 2.228 \times 3.8 \times 10^{-5} = (0.00106, 0.00122)$, o sea, el valor de la pendiente β está entre 0.00106 y 0.00122 de lo cual se tiene una confiabilidad del 95%.

Un intervalo de confianza para la media $\mu_{y.x}$ puede calcularse mediante:

$$\hat{y}_j \pm t_{\alpha/2,n-2} S\left(\hat{y}_j\right). \tag{11.12}$$

La varianza del estimador es:

$$S^2\left(\hat{y}_j\right) = MCE\left[1/n + (1/S_{xx})\ (x_j - \bar{x})^2\right].$$

$$= 0.00033\ [1 / 12 + (310 - 300)^2 / 228\,800] = 2.76 \times 10^{-5}.$$

Un intervalo de 95% de confianza para $\mu_{y.x}$ cuando la concentración x_j es 310 ppm, se calcula como:

$$0.342 \pm 2.228 \times 0.00526 = (0.33, 0.35) \text{ en unidades de DO.}$$

La amplitud del intervalo depende de la distancia $x_j - \bar{x}$: mientras más lejos de la media se elija el punto x_j, más amplio será el intervalo de confianza. Estos intervalos forman una región de confianza alrededor de la recta de regresión limitada por una hipérbola, como se muestra en la figura 11.3.

11.4 Análisis de los residuos

La regresión sirve para describir la relación funcional, mediante una ecuación matemática, entre variables dependientes e independientes. Por ejemplo, se usa para hacer pronósticos, es decir, para predecir valores de Y dado un valor de X. También tiene aplicación en el control de procesos industriales, en la calibración de instrumentos, en la descripción del crecimiento biológico y en la estimación de probabilidades de eventos. En estas situaciones, y en cualquier otra aplicación, se requiere que el modelo cumpla las condiciones de validez y de buen ajuste.

Para que tenga validez, el modelo debe cumplir los supuestos iniciales de independencia y normalidad de los residuos, además de la homogeneidad de las varianzas de los valores de Y para cada valor de X. Si se trata del modelo (11.3), el diagrama de puntos debe mostrar una relación lineal entre Y y X. Cuando no se tienen repeticiones de Y para cada X_i, los supuestos de normalidad y de homogeneidad de varianzas no pueden probarse en el sentido teórico. Sin embargo, pueden hacerse chequeos

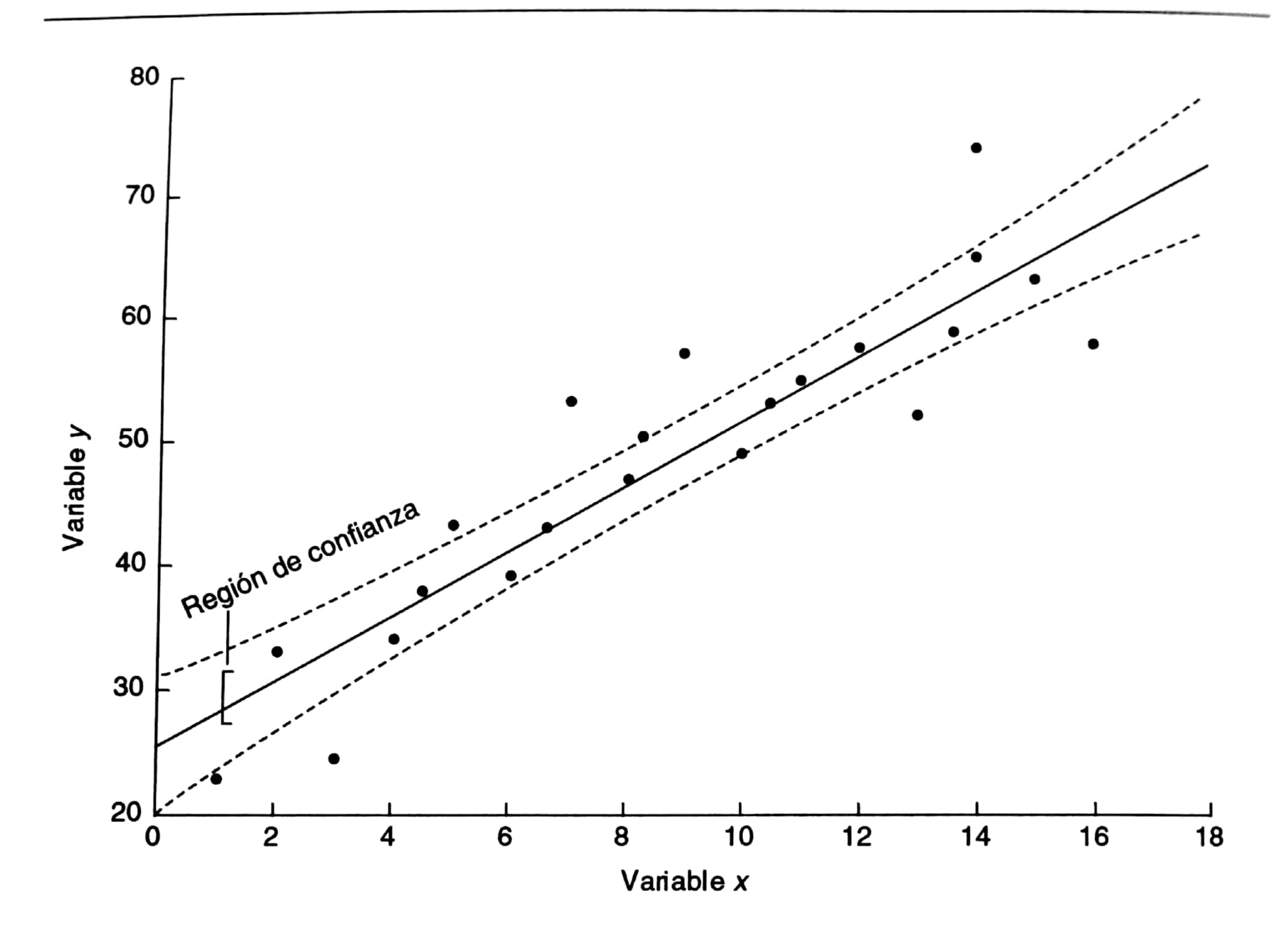


Figura 11.3 Región de confianza para $E(Y/X) = \mu_{y.x}$

razonables con los residuos obtenidos del modelo ajustado. Los gráficos de residuos versus X o versus $\hat{y}$ y de probabilidad normal, se construyen e interpretan como se hizo en el capítulo 5. En las figuras 11.4 y 11.5 se muestran el gráfico de residuos y el de ajuste a la normal para los datos del ejemplo 11.1.

Se deduce de estos gráficos que los residuos son independientes —banda de puntos paralela al eje horizontal— y tienen distribución Normal. Igualmente útil es el gráfico de residuos e_i contra $\hat{y}_i$, este reemplaza el anterior, sobre todo cuando se tienen pocos valores de X. Una discusión más amplia sobre análisis de residuos en regresión puede leerse en Hocking (1996), Neter y otros (1996), o Draper y Smith (1998).

Cuando no se cumplen los supuestos, puede ensayarse una transformación de la variable Y. Las transformaciones logarítmica e inversa permiten linealizar la función, mientras que la raíz cuadrada estabiliza las varianzas. En algunos casos es necesario transformar ambas variables; en otros es suficiente con la transformación de X. Cuando no se tiene éxito con ninguna transformación, lo recomendable es abandonar el modelo propuesto y ensayar otros modelos, tal vez más complejos, como los mínimos cuadrados ponderados si la varianza de Y crece al aumentar X, o modelos polinomiales o exponenciales. Otra opción incluye los modelos lineales generalizados que permiten distribuciones no normales de los errores ε_i, en

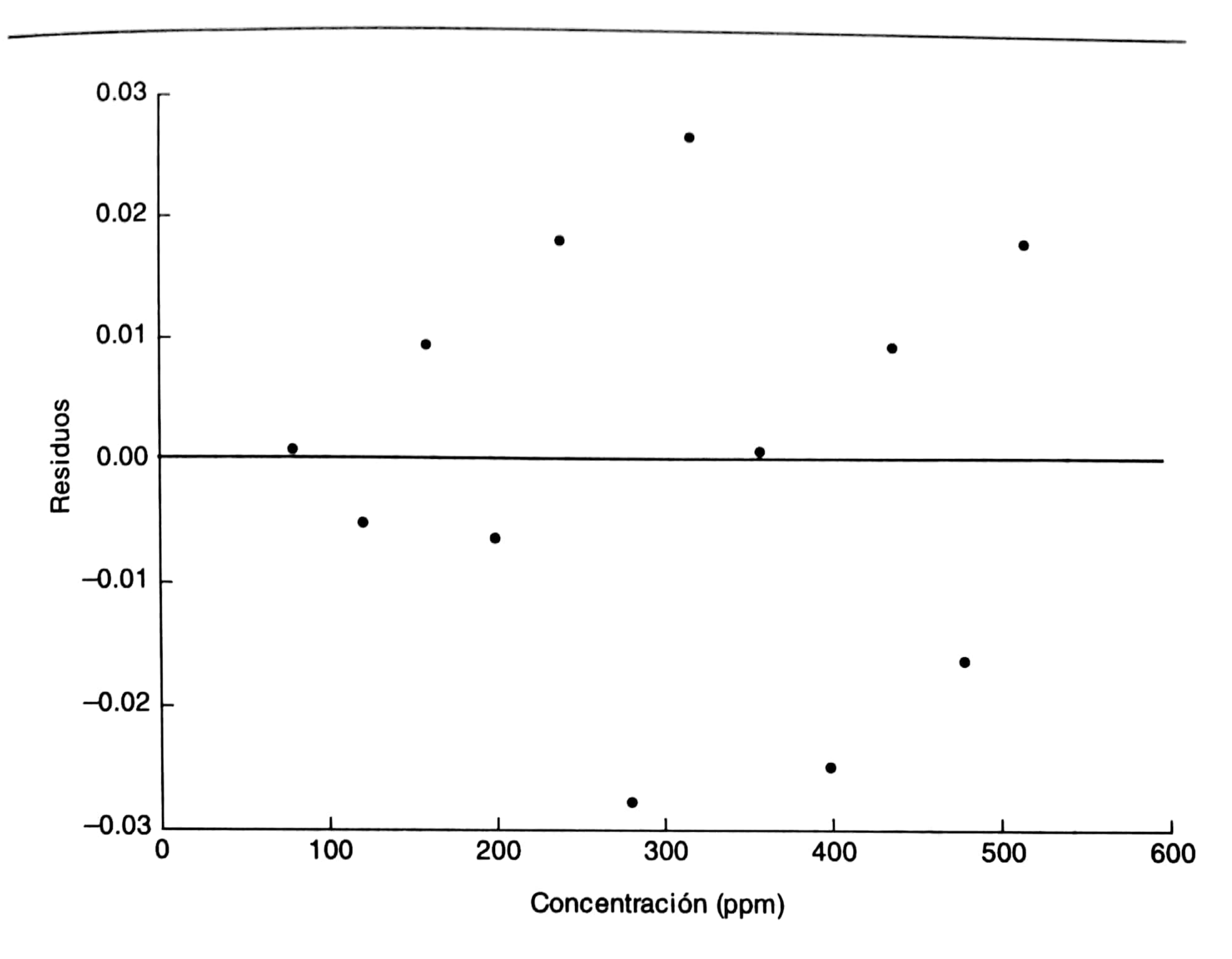


Figura 11.4 Gráfico de residuos versus *X* para el ejemplo 11.1

particular, los modelos aditivos generalizados, Hastie y Tibshirani (1999), Quinn y Keough (2003).

La no independencia en los errores del modelo lineal se reconoce por la presencia de correlaciones, temporales o espaciales, entre residuos adyacentes. Para estos modelos autocorrelacionados se dispone de metodologías específicas como los modelo para datos longitudinales y para series cronológicas.

11.5 Falta de ajuste y diagnóstico en regresión

Para utilizar el análisis de regresión en una investigación, se requiere proponer un modelo y ajustar una línea de regresión a dicho modelo. Este proceso puede llevarse a cabo en uno o varios pasos, es decir, no siempre el primer ajuste es el que se acepta. Para juzgar si el modelo es correcto, o sea, si tiene buen ajuste, se requiere considerar, desde el diseño inicial de la investigación, la posibilidad de obtener varios valores de Y para cada nivel de la variable X.

En el caso de la regresión lineal simple, la falta de ajuste al modelo puede detectarse examinando el diagrama de puntos. Cuando el modelo es más complejo o involucra varias variables, el análisis gráfico no es posible.

Cuando se tienen valores repetidos de Y para cada X_i, es posible subdividir la *SCE*

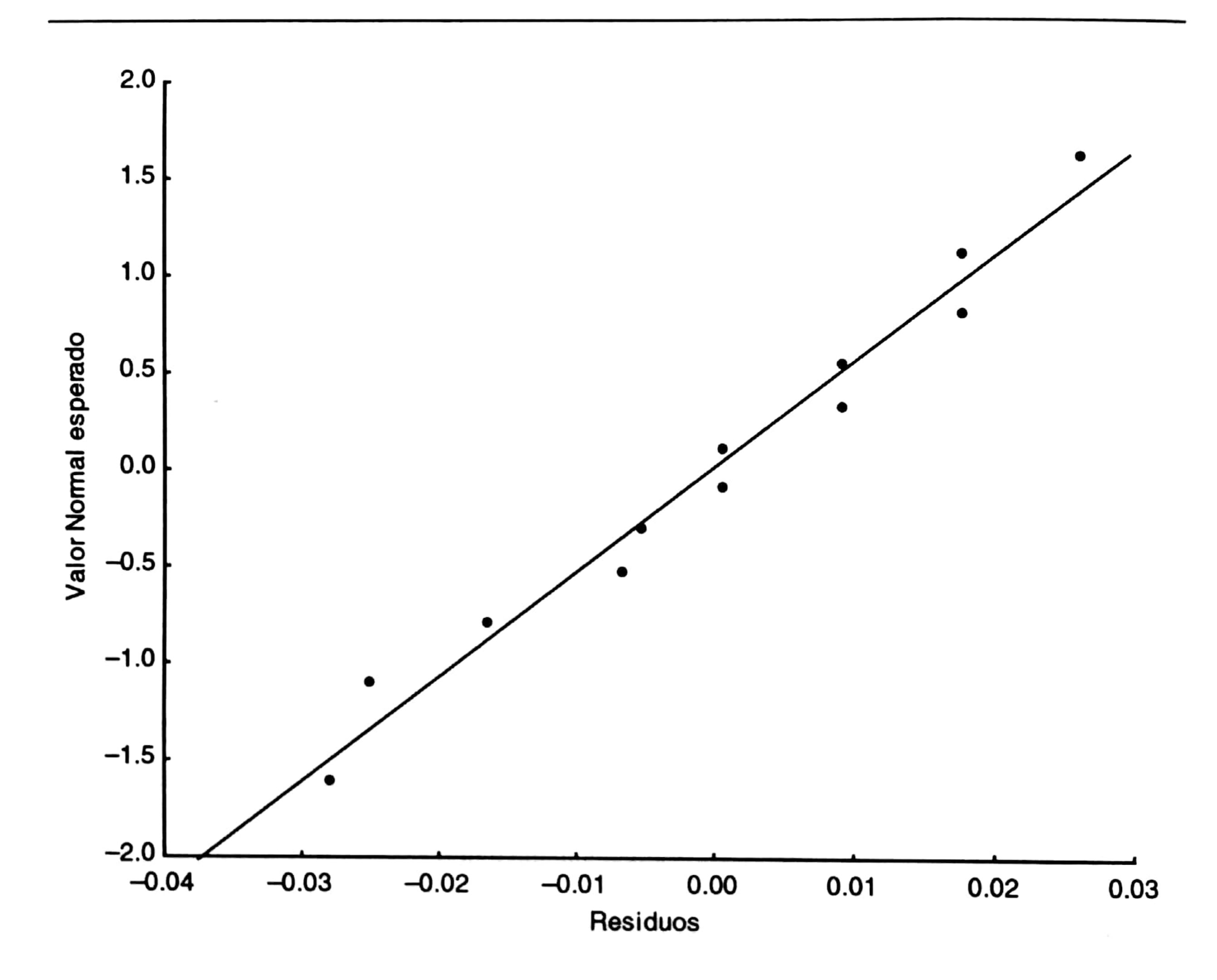


Figura 11.5 Gráfico de probabilidad normal para los residuos del ejemplo 11.1

en dos sumas: una correspondiente al error puro, *SCEP*, y la otra asignada a la falta de ajuste, *SCFA*. Entonces *SCE* = *SCEP* + *SCFA*. La *SCEP* se calcula como la suma de las diferentes sumas de cuadrados de los valores de *Y* para cada valor de *X*. Supóngase *m* valores diferentes de *X*, cada uno con r_i, valores de *Y*. Entonces,

$$SCEP = \sum_{i=1}^{m}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_{j.}\right)^2; \text{ con } n_e = \sum_i r_i - m \ gl.$$

Para probar la falta de ajuste, se completa la tabla 11.1 del ANOVA con las fuentes de variación que se muestran en la tabla 11.3.

Los pasos recomendados para efectuar el análisis son:

1. Se ajusta el modelo y se construye la tabla del ANOVA, pero no se realiza la prueba *F* para regresión.

2. Se calcula la *SCEP* y se realiza la prueba *F* para la falta de ajuste, como se indica en la tabla 11.3 (H_0: no hay falta de ajuste).

3. Si no hay significancia, se efectúa la prueba *F* para la regresión y se continúa con el análisis.

4. Si la falta de ajuste es significativa ($p < 0.05$), no se continúa con el análisis

Tabla 11.3 Prueba F para falta de ajuste

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Falta de ajuste	*SCFA*	$(n-2)-n_e$	MCFA	*MCFA* / MCEP	0.*xx*
Error puro	*SCEP*	n_e	*MCEP*		

de regresión. Se busca, en cambio, mejorar el modelo; para hacerlo, lo recomendable es apoyarse en el análisis de los residuos.

Ejemplo 11.2 Para estudiar el crecimiento de los fetos en madres de la población de Medellín (Colombia), José Roberto Gómez M. midió, por medio de exámenes ecográficos, el área cefálica desde la semana 12 hasta la semana 41 del embarazo. Los datos proporcionados por el autor se muestran en la tabla 11.4.

La variable dependiente Y es el área cefálica y el tiempo de gestación en semanas es la variable independiente X. Entonces, por cada valor de X se tienen cinco repeticiones de Y. La ecuación lineal tiene la forma:

$\hat{y}_i = -33.97 + 3.04\, x_i$ con $R^2 = 98.3\%$ y $S_b = 0.065$.

La tabla 11.5 del ANOVA permite probar la falta de ajuste al modelo lineal y las sumas de cuadrados se calculan como en los ejemplos anteriores, aunque en este caso es recomendable el uso del computador.

Para calcular la *SCEP* es necesario calcular por aparte la suma de cuadrados para cada valor de X, así:

Para $X = 12$, $S_{12} = \sum y_i^2 - (1/r_i)(\sum y_i)^2 = 110.217 - (1/5)(23.26)^2 = 2.0115$.

Para $X = 16$, $S_{16} = 1096.904 - (1/5)(73.04)^2 = 29.9357$, y en forma similar para los otros valores de X. Los resultados son:

$S_{12} = 2.0115$.
$S_{16} = 29.9357$.
$S_{20} = 24.8793$.
$S_{24} = 38.5325$.
$S_{28} = 34.5057$.
$S_{32} = 107.6399$.
$S_{36} = 98.6854$.
$S_{40} = 46.0794$.
Suma = 382.2698 = *SCEP*.

Tabla 11.4 Área cefálica (en cm^2) de fetos desde la semana 12 hasta la semana 40 del embarazo en cinco madres de la ciudad de Medellín

Madres	*Semanas*							
	12	*16*	*20*	*24*	*28*	*32*	*36*	*40*
1	4.28	12.37	23.16	*37.77*	56.84	65.04	69.80	85.32
2	4.66	19.35	27.36	36.03	55.03	67.42	72.09	83.89
3	4.83	13.43	20.49	40.43	50.45	54.52	70.49	91.59
4	5.70	13.65	22.61	39.30	52.59	65.97	80.76	90.82
5	3.79	14.24	23.16	32.51	57.45	65.31	78.46	89.13

Tabla 11.5 Tabla del ANOVA para los datos del ejemplo 11.4

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Debido a la regresión	31 037.1	1	31 037.1	2 170.4	<0.001
Error	545.1	38	14.3		
Total	31 582.2	39			
Falta de ajuste	162.9	6	27.15	2.27	0.061
Error puro	382.2	32	11.94		

Entonces $SCFA = SCE - SCEP = 545.1642 - 382.2698 = 162.8944$, y los *gl* son $n_e = 40 - 8 = 32$ para la *SCEP* y $40 - 2 - 32 = 6$ para la *SCFA*.

Dado que no hay falta de ajuste ($p = 0.061$) y la regresión es altamente significativa ($p < 0.001$), se concluye que el área cefálica crece en forma lineal con el tiempo. Además, los supuestos de normalidad e independencia de los residuos y de homogeneidad de varianzas son todos aceptables.

Por otra parte, como la ecuación de regresión es un modelo para hacer pronósticos, es posible hallar el área cefálica promedio para una semana cualquiera en el periodo de la semana 12 a la 40 de gestación. Por ejemplo, en la semana 25 se espera que el área cefálica promedio sea de $\hat{y}_{25} = -33.97 + 3.04 \times 25 = 42.03$ cm^2.

Un intervalo de confianza del 95% para $\mu_{y.x}$ en $X = 25$, se calcula usando la expresión (11.12). El error estándar se obtiene de $S^2(\hat{y}_{25}) = 545.1\left[(1/40) + (25 - 26)^2/3\,358.4\right] = 13.789$, $S(\hat{y}_{25}) = 3.71$. Luego, el intervalo es $42.03 \pm 2.025 \times 3.71 = (34.52, 49.54)$ cm^2.

En investigaciones como la anterior puede ser de interés el problema inverso llamado regresión inversa: si se conoce el área cefálica averiguar el tiempo probable de embarazo. Sea el área cefálica conocida $y_0 = 62.50$ cm^2, entonces la semana más probable de gestación será:

$$\hat{X}_0 = (y_0 - a)/b = (62.50 + 33.97)/3.04 = 31.7.$$

Esto es, aproximadamente la semana 32.

Los límites de confianza inversos para X_0 son más complejos de calcular. Las fórmulas respectivas se pueden consultar en los libros de Sokal y Rohlf (1995) o de Draper y Smith (1998).

El diagnóstico se refiere a la idoneidad del modelo, o sea, qué tan bien se ajusta el modelo a los datos de la muestra. ¿Es apropiado el modelo lineal o debería cambiarse por uno curvilíneo? ¿Hay datos extremos que influyen atípicamente en la estimación de los parámetros? La influencia puede presentarse porque el dato está en los extremos del rango de X o porque tiene un residuo grande. Hay varias medidas de diagnóstico que permiten identificar datos influyentes y detectar no linealidad, Quinn y Keough (2003).

El *leverage* (influjo o potencia) es una medida de la condición extrema de una observación con respecto a la variable X. Una observación con alto *leverage* es un *outlier* en el espacio de X y mide, básicamente, la influencia de x_i sobre $\hat{y}_i$. El *leverage*, escrito h_i, varía entre 1 y $1/n$ y un dato con *leverage* mayor que $4/n$ es sospechoso y digno de analizarse.

Otra medida diagnóstica es la *influencia* de cada punto. Esta mide el efecto que tiene una observación sobre la línea de regresión ajustada. Se estima mediante la distancia D_i de Cook y refleja la influencia que tiene esta observación sobre la pendiente de la recta. Una D_i mayor que el resto indica que el dato es influyente, es decir, que al eliminarlo del análisis, la pendiente puede cambiar significativamente.

Los residuos estudentizados de cada observación, definidos $e_{si} = e_i \Big/ \sqrt{MCE\,(1 - h_i)}$ son otro elemento para el diagnóstico. Estos se comparan entre sí mediante la distribución t de Student y los que resulten significativamente grandes corresponder a observaciones

que pueden ser *outliers*. Tanto el diagnóstico gráfico como el numérico requieren un buen *software* estadístico para su realización.

11.6 Transformaciones en regresión

Cuando los supuestos del modelo lineal no se satisfacen, una alternativa es cambiar de modelo. En general, el nuevo modelo es más complejo, sin embargo en muchos estudios puede convertirse en uno lineal, aplicando una transformación logarítmica, de raíz cuadrada, recíproca, etc., a los datos.

Cuando la ecuación lineal no es apropiada para expresar la relación entre las variables, deberá ensayarse un modelo curvilíneo o no lineal. Para escoger la forma de la curva es importante conocer la literatura sobre el campo de investigación, y también ensayar algunos modelos que muestren buen ajuste y cumplan los supuestos probabilísticos iniciales. Algunos modelos no lineales en los parámetros son intrínsecamente lineales, ya que se pueden linealizar mediante una transformación de las variables. En la siguiente tabla se muestran varios ejemplos, véase también Hocking (1996).

Ejemplo 11.3 El periodo de entrenamiento influye en el desempeño de muchas actividades. Midiendo el tiempo de entrenamiento en meses y el éxito en la realización de una actividad en puntajes sobre una escala apropiada, se tienen los siguientes datos:

Meses (X):	1	1	2	2	3	3	4	5	5	5
Puntaje (Y):	45	40	51	53	76	70	115	150	145	148

El diagrama de puntos (véase figura 11.6) muestra una relación curvilínea entre las dos variables. Se ensaya el segundo modelo de la tabla anterior y se ajusta una recta a los puntos $(x_i, \log_{10} y_i)$; el nuevo diagrama de puntos se ve en la figura 11.7.

Los logaritmos de los puntajes se muestran en la tabla 11.7 y en la 11.8 aparece el ANOVA para la regresión.

Las sumas requeridas son:

$$S_{xy} = 62.07810 - (31 / 10) \times 18.90145 = 3.23561.$$
$$S_{xx} = 22.9 \text{ y } S_{yy} = 0.46431.$$

Las estimaciones de los coeficientes del modelo transformado son:

Tabla 11.6 Ejemplos de transformaciones de variables que linealizan el modelo de regresión

Modelo original	*Modelo transformado*
1. $Y = \alpha X^{\beta}$	$\log Y = \log \alpha + \beta \log X$
2. $Y = \alpha \beta^{X}$	$\log Y = \log \alpha + X \log \beta$
3. $e^{Y} = \alpha X^{\beta}$	$Y = \ln \alpha + \beta \ln X$
4. $Y = \alpha + \beta / X$	$Y = \alpha + \beta X'$, con $X' = 1 / X$
5. $Y = 1 / (\alpha + \beta X)$	$Y' = \alpha + \beta X$, con $Y' = 1 / Y$

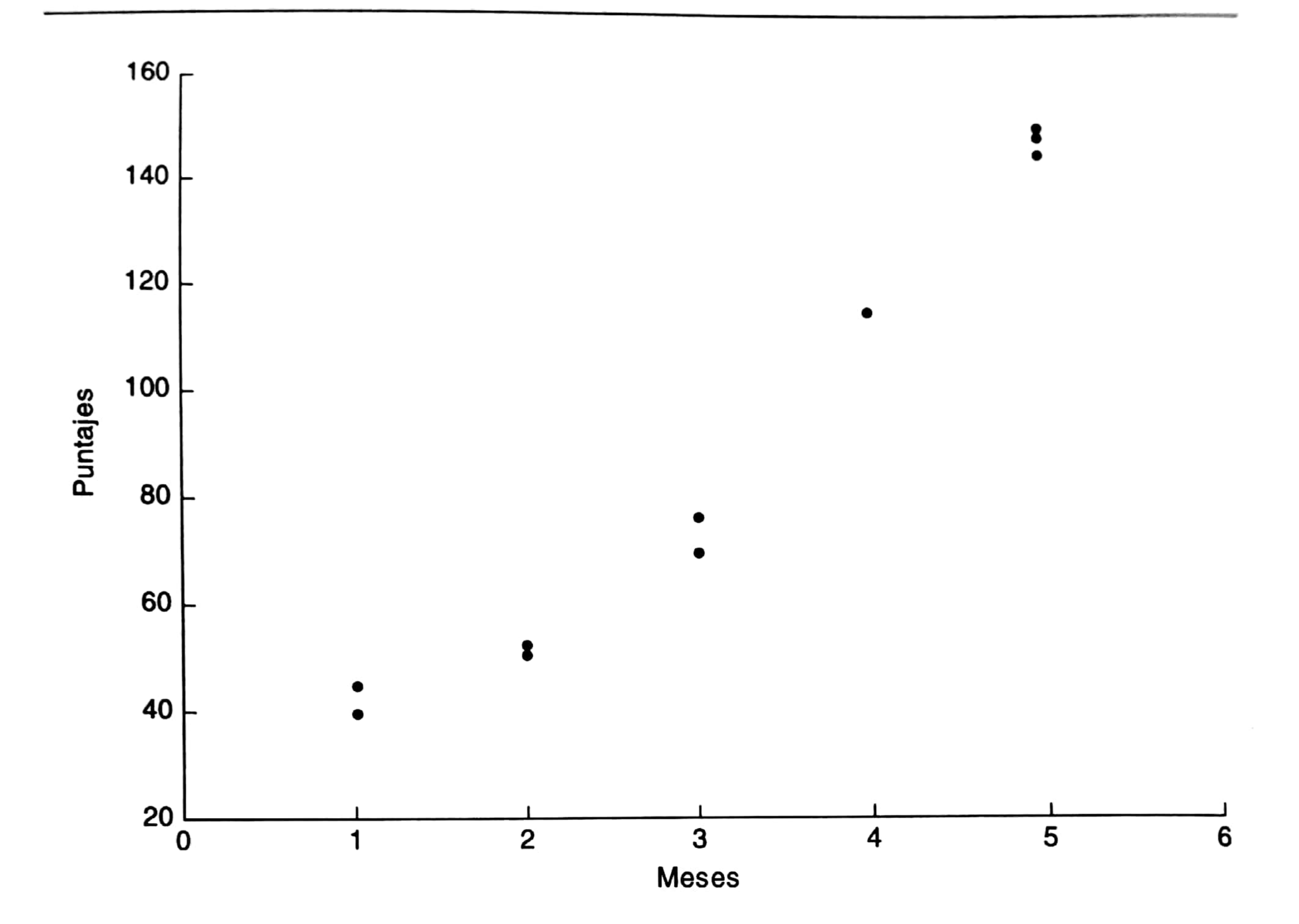


Figura 11.6 Diagrama de puntos para el ejemplo 11.3

Tabla 11.7 Logaritmos de los puntajes para el ejemplo 11.3

	X_i	Y_i	Log_{10} Y_i
	1	45	1.65321
	1	40	1.60206
	2	51	1.70757
	2	53	1.72428
	3	76	1.88081
	3	70	1.84510
	4	115	2.06070
	5	150	2.17609
	5	145	2.16137
	5	148	2.17026
Suma	31		18.98145

$\log b = S_{xy} / S_{xx} = 0.14129.$

$\log a = 1.89014 - 0.14129 \times 3.1 = 1.45214.$

La ecuación logarítmica ajustada es, entonces:

$$\log \hat{y}_i = 1.45214 + 0.14129x_i,\ R^2 = 0.9846.$$

El modelo original es $\mu_{y.x} = \alpha\beta^x$. Las estimaciones de los parámetros son:

$a = \text{antilog}(1.45214) = 28.32.$

$b = \text{antilog}(0.14129) = 1.38.$

Entonces el modelo original ajustado será $\hat{y}_i = (28.32)(1.38)^x$. Si el entrenamiento dura seis meses, se espera un puntaje promedio igual a $\hat{y}_6 = (28.32)(1.38)^6 = 196$. Aquí se está extrapolando para un valor de X por fuera de su rango, por lo que el resultado obtenido puede ser bastante impreciso.

Un intervalo del 95% de confianza para $\mu_{y.x}$ tiene la forma (11.12) calculado con los datos transformados. Entonces,

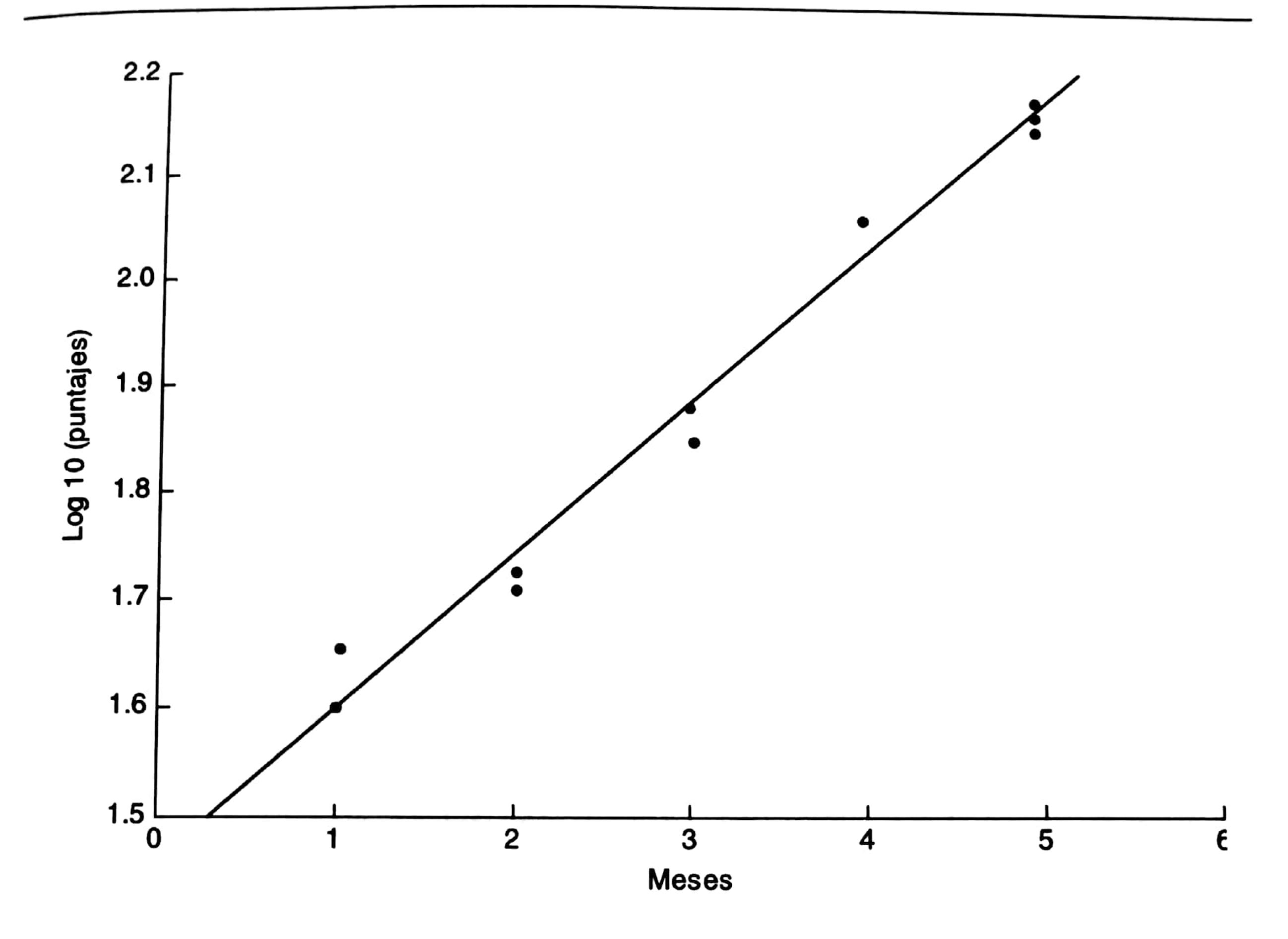


Figura 11.7 Datos del ejemplo 11.3 con puntajes transformados en logaritmos

$S^2(\hat{y}) = 0.00089\left[(1/10) + (6-3.1)^2/22.9\right] = 0.000416,\ S(\hat{y}) = 0.02040.$

$\log \hat{y}_6 = 1.45214 + 0.14129 \times 6 = 2.29988.$

Por tanto, el intervalo para la media es:

$2.29988 \pm 2.306 \times 0.02040 = (2.25285, 2.34691).$

Extrayendo el antilogaritmo a estos límites se obtiene el intervalo (179, 222) para el puntaje promedio cuando el entrenamiento es de seis meses.

Ejemplo 11.4 Para estudiar la morfología de *Micrathyria ocellata,* Ortega (1992) midió el largo del cuerpo, LC, y el ancho de la cápsula cefálica, ACC, de varios ejemplares en cada uno de sus 11 instares. A continuación se listan los resultados de los valores promedio por instar, y en la tabla 11.9 se ilustra el ANOVA para la regresión de este ejemplo.

Tabla 11.8 Tabla del ANOVA para la regresión del ejemplo 11.3

Fuente de variación	*SC*	*gl*	*MC*	*F*
Debida a la regresión	0.45717	1	0.45717	512.2
Error	0.00714	8	0.00089	
Total	0.46431	9		

LC (mm) X:	1.17	1.40	1.72	2.31	3.05	4.11	5.58	7.68	10.13	13.24	16.85
ACC (mm) Y:	0.32	0.41	0.53	0.74	0.98	1.29	1.70	2.24	2.97	3.63	4.58

Se desea conocer el crecimiento relativo de una parte del cuerpo con respecto a la longitud total del mismo; para ello se utiliza la fórmula propuesta por Huxley que tiene la forma $Y = kX^{\beta}$, donde el exponente β representa la constante de crecimiento. Si $\beta = 1$, el crecimiento es isométrico y si $\beta \neq 1$, el crecimiento es alométrico, es decir, la parte en cuestión crece en diferente proporción al cuerpo.

Utilizando la transformación logarítmica se linealiza la ecuación, con lo cual el coeficiente β resulta igual a la pendiente de la recta. Para hallar β se ajustan los logaritmos de los datos al modelo $\log Y = \log k + \beta \log X$. Aquí los cálculos son dispendiosos y se recomienda usar el computador; si se usa la calculadora de bolsillo, es conveniente trabajar con un mínimo de cinco decimales para evitar errores de redondeo.

La ecuación de regresión ajustada tiene la siguiente forma:

$$\hat{y}' = -0.51245 + 0.97521x',\ R^2 = 0.9959.$$

El ajuste al modelo linealizado es bueno, lo que permite concluir que la pendiente b estima correctamente la constante de crecimiento β.

Ahora se deben probar las hipótesis H_0: $\beta = 1$ contra H_1: $\beta \neq 1$. Si los supuestos del modelo son válidos, el cociente $(b - b_0) / S_b$ tiene la distribución t de Student con $n - 2$ *gl*. El error estándar se calcula como en el ejemplo 11.2 y se obtiene el valor $S_b = 0.02092$. La prueba exige comparar la pendiente β con la unidad, o sea con $b_0 = 1$. Entonces se calcula $T = (0.97521 - 1) / 0.02092 = -1.18$.

Para un nivel de significación $\alpha = 0.05$, el valor tabulado es $t_{0.025,9} = 2.26$. Puesto que $|T| < 2.26$, no se rechaza H_0 y se concluye que el crecimiento de la cápsula cefálica es isométrico respecto a la longitud del cuerpo.

La ecuación inicial ajustada se obtiene reemplazando la constante k por anti-log(−0.51245) = 0.307. La fórmula de Huxley será $\hat{y} = 0.307x^{0.975}$. Por ejemplo, si el largo corporal es $x = 8.70$ mm, el ancho promedio de la cápsula cefálica se estimará en $\hat{y} = 0.307(8.7)^{0.975} = 2.53$ mm.

Muchas relaciones entre variables son no lineales por naturaleza y las transformaciones usuales no las hacen lineales. Una manera de aproximarse al modelo no lineal es ajustar un polinomio en X de segundo grado o mayor según la precisión deseada. Dicho proceso exige demasiados cálculos aritméticos por lo que, normalmente, deberá usarse el computador. Para un ejemplo suficientemente ilustrativo del proceso de ajuste polinomial véase Zar (1999).

11.7 Variable independiente aleatoria

En las secciones anteriores se ha supuesto que la variable independiente X es controlada por el investigador, o que no está sujeta a variación aleatoria; este es el modelo I de regresión. En la práctica dicho supuesto no es totalmente válido; por ejemplo, en estudios observacionales es común medir ambas variables en la misma unidad de estu-

Tabla 11.9 Tabla del ANOVA para la regresión del ejemplo 11.4

Fuente de variación	*SC*	*gl*	*MC*	*F*
Debida a la regresión	1.53589	1	1.53589	2163.2
Error	0.00636	9	0.00071	
Total	1.54225	10		

dio, de modo que la variable independiente está sujeta no solo a la variabilidad de los errores de medida sino también, y fundamentalmente, a la variación entre las unidades en la población; este es el modelo II de regresión. La talla y el peso entre individuos son ejemplos de variables de regresión donde ambas son aleatorias. Considerando la talla como variable independiente, se observa que está sujeta a errores de medida y a la variación entre los individuos. Cuando tanto X como Y son variables aleatorias, el análisis de regresión resulta complejo, aun en el caso de regresión simple.

Según Draper y Smith (1998), puede aplicarse el análisis de regresión usual, o sea basado en mínimos cuadrados, cuando el rango de valores de X es grande comparado con su variabilidad aleatoria o cuando el objetivo principal de la regresión es la predicción. Como recomendación práctica debe buscarse, en lo posible, que los valores de X caigan en un rango suficientemente amplio, de modo que al compararse con la variabilidad aleatoria, esta resulte pequeña y pueda ignorarse.

Ejemplo 11.5 Pareja (1985) estudió la ecología y el comportamiento de arañas *(Diplura* sp.*)* en varios bosques de Caucasia, departamento de Antioquia (Colombia), y midió, entre otras variables, su peso corporal y el ancho de sus redes. Algunos de sus datos aparecen listados en la tabla 11.10.

Ambas variables requieren medición y varían con cada unidad de la muestra; dado que el peso corporal es medido con suficiente precisión, se espera que su variabilidad sea pequeña comparada con el rango de los pesos.

El diagrama de puntos (véase figura 11.8) presenta bastante dispersión, pero al aumentar el peso, el ancho de la red tiende a ser mayor, así que se puede ensayar un modelo lineal.

La ecuación de regresión ajustada es $\hat{y}_i = 25.91 + 15.37x_i$. Aunque el ANOVA muestra que la regresión es significativa ($p = 0.003$), el ajuste no es bueno. El peso de la araña explica el 23.6% de la variación en el ancho de la red; el 76.4% restante se debe a otras causas o variables no tenidas en cuenta en la regresión. Pudo ocurrir que no se eligió la variable explicativa correcta, o que existen muchas variables que influyen simultáneamente en el tamaño de la red y solo se analizó una de ellas.

De otro lado, el error estándar S_b es grande comparado con el valor de b, lo que explica, en parte, la imprecisión en el ajuste lineal —la pendiente de la recta es muy variable—.

Tabla 11.10 Peso corporal X (g) y ancho de la red Y (cm) para 36 arañas

X	Y	X	Y	X	Y
1.64	*70*	2.20	60	0.80	70
1.70	75	2.00	25	0.18	20
1.55	45	0.92	20	0.18	15
1.65	50	1.45	20	0.12	30
2.90	120	1.00	50	0.05	4
1.70	50	1.30	80	0.11	5
2.20	10	1.70	50	0.45	20
0.05	50	2.00	80	0.40	43
1.84	50	0.16	80	2.96	100
0.30	20	0.16	20	1.60	60
1.50	10	1.70	20	0.05	36
1.72	45	1.70	30	0.05	36

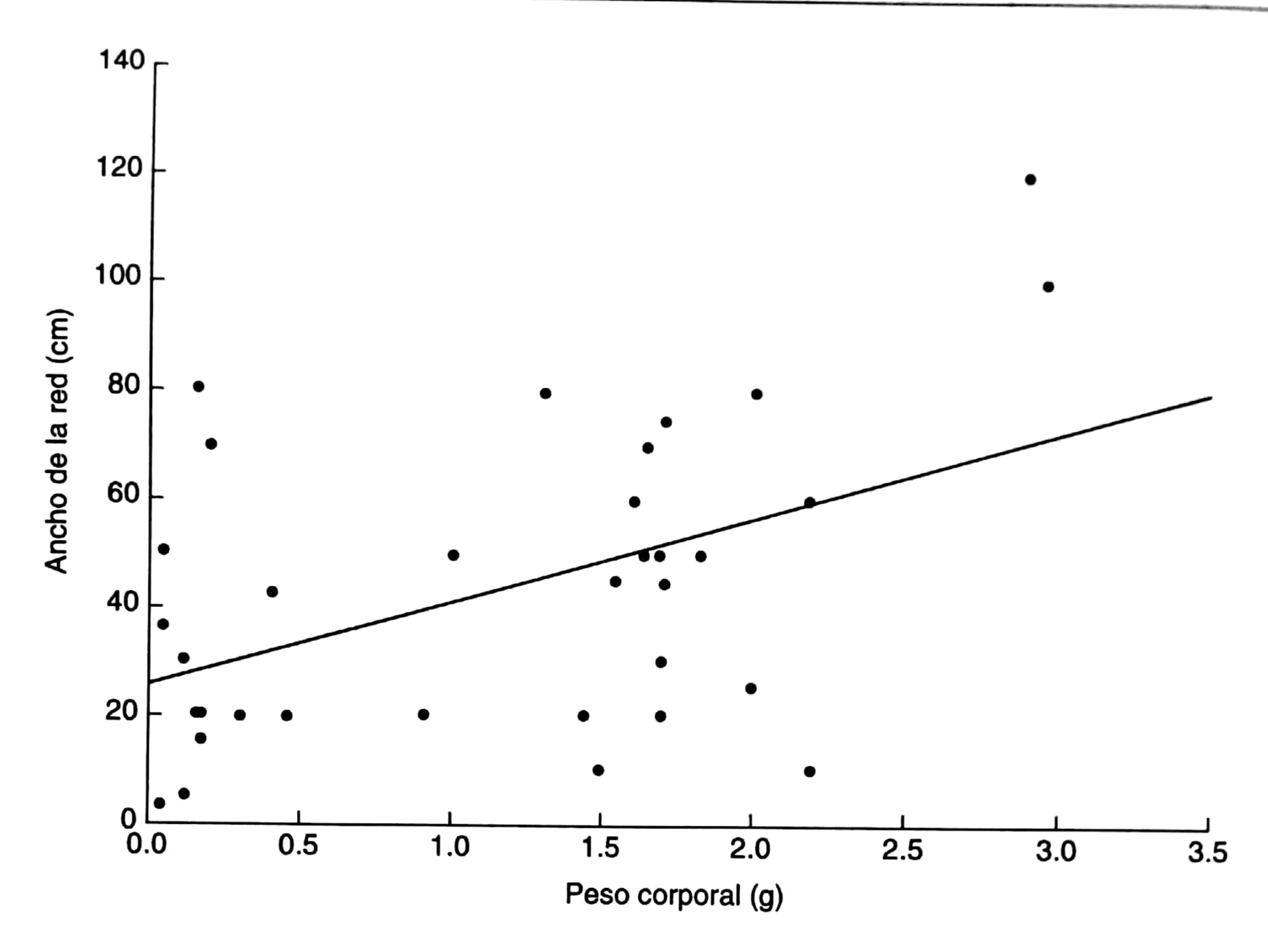


Figura 11.8 Diagrama de puntos para el ejemplo 11.5

En efecto, $S_{xx} = SCR / b^2 = 26.954$, luego $S_b = \sqrt{MCE / S_{xx}} = 4.75$. y el intervalo del 95% de confianza para β es $15.37 \pm 2.034 \times 4.75 = (5.71\ , 25.03)$ que resulta demasiado amplio.

En cuanto a los supuestos, tanto el gráfico de residuos contra X como el de ajuste a la normal, muestran que dichas condiciones se satisfacen plenamente.

En este ejemplo de regresión lineal simple, con la variable X aleatoria, se encontró alta imprecisión tanto para los valores ajustados como para la pendiente. Se puede afirmar que el modelo no describe adecuadamente la relación funcional entre el ancho de la red y el peso corporal de las arañas. Gráficamente no se observa un modelo mejor para establecer la relación por lo que sería aconsejable analizar estos datos con metodologías más recientes.

11.8 Otros métodos de regresión

Desarrollos recientes analizan el modelo II de regresión de acuerdo con las inten-

Tabla 11.11 Tabla del ANOVA para la regresión del ejemplo 11.5

Fuente de variación	*SC*	*gl*	*MC*	*F*	*p*
Debida a la regresión	6367.6	1	6367.6	10.5	0.003
Error	20657.2	34	607.6		
Total	27024.8	35			

ciones de la investigación. Si el objetivo principal es la predicción, puede usarse el modelo I aunque *X* sea aleatoria, con los demás supuestos básicos de este modelo.

Por otra parte, si se pretende una descripción de la relación de *Y* con *X*, no es recomendable el método de los mínimos cuadrados porque la variabilidad del error asociado tanto a *X* como a *Y* genera estimaciones sesgadas de la pendiente β.

Entre los métodos propuestos para reemplazar el de los mínimos cuadrados cabe mencionar el de la *regresión del eje mayor* MA *(major axis)* y el de la *regresión del eje mayor reducido o estándar* RMA *(reduced major axis)*, Legendre y Legendre (1998).

Teniendo en cuenta que la regresión de *Y* sobre *X* genera una línea de regresión diferente de la regresión de *X* sobre *Y* para los mismos datos (*X* y *Y* aleatorias), se observa que la recta obtenida mediante la regresión RMA es intermedia entre las dos anteriores y, por tanto, intuitivamente parece la de mejor ajuste. Bajo ciertas condiciones de proporcionalidad entre los errores de medición de *X* y *Y*, se sugiere escoger el método de regresión RMA si la varianza σ_x^2 es mayor que un tercio de σ_y^2, en otro caso, el modelo I es aceptable (Quinn y Kenough, 2003).

El modelo II de regresión, igual que la regresión ponderada, ha tenido poca acogida en la literatura biológica, al parecer por el desconocimiento del riesgo potencial en las estimaciones mínimo cuadráticas cuando *X* es aleatoria. Además, todas estas metodologías exigen programas específicos de computador, desarrolladas paralelamente con ellas, pero no siempre disponibles en el momento de realizar el análisis de los datos.

El método de los mínimos cuadrados en regresión presupone distribución normal de los errores y es sensible a observaciones influyentes o extremas.

Gracias a la capacidad y velocidad cada vez mayor de los computadores, fue posible enriquecer el modelo lineal con nuevos métodos que tienen menos supuestos y potencialmente descubren más propiedades en un conjunto de datos.

Entre los nuevos procedimientos para ajustar modelos de regresión se tienen los siguientes:

a. Las técnicas de regresión robusta que minimizan la influencia tanto de valores extremos como de la no normalidad de los errores. Entre ellas pueden citarse el método de las *mínimas desviaciones absolutas*, el método de los *estimadores M de Huber* y la *regresión no paramétrica* (basada en rangos) (Birkes y Dodge, 1993).

b. Los métodos de suavizamiento *(smoothing)* que no presuponen ningún modelo (lineal o no lineal) y que describen la relación entre *Y* y *X* con mínimas restricciones sobre la forma que puede tomar dicha relación. Un ejemplo es el suavizamiento en estudios cronológicos llamado de *promedios o medianas móviles*. Más recientemente se han propuesto varios métodos de suavizamiento de nubes de puntos por segmentos que luego se unen formando una curva suave, uno de ellos es el método de *suavizamiento localmente ponderado* (Lowess), conocido también como *regresión robusta ponderada localmente* (Goodall, 1990; Simonoff, 1996). Una aplicación de dicho método para los datos del ejemplo 11.5 se muestra en la figura 11.9.

c. Otro de dichos métodos es el *modelo aditivo generalizado* que consiste en una suma de funciones suavizadas, es no paramétrico en cuanto a que no impone una forma paramétrica a las funciones sino que estas se estiman mediante iteraciones utilizando suavizadores de nubes de puntos, Hastie y Tibshirani (1999).

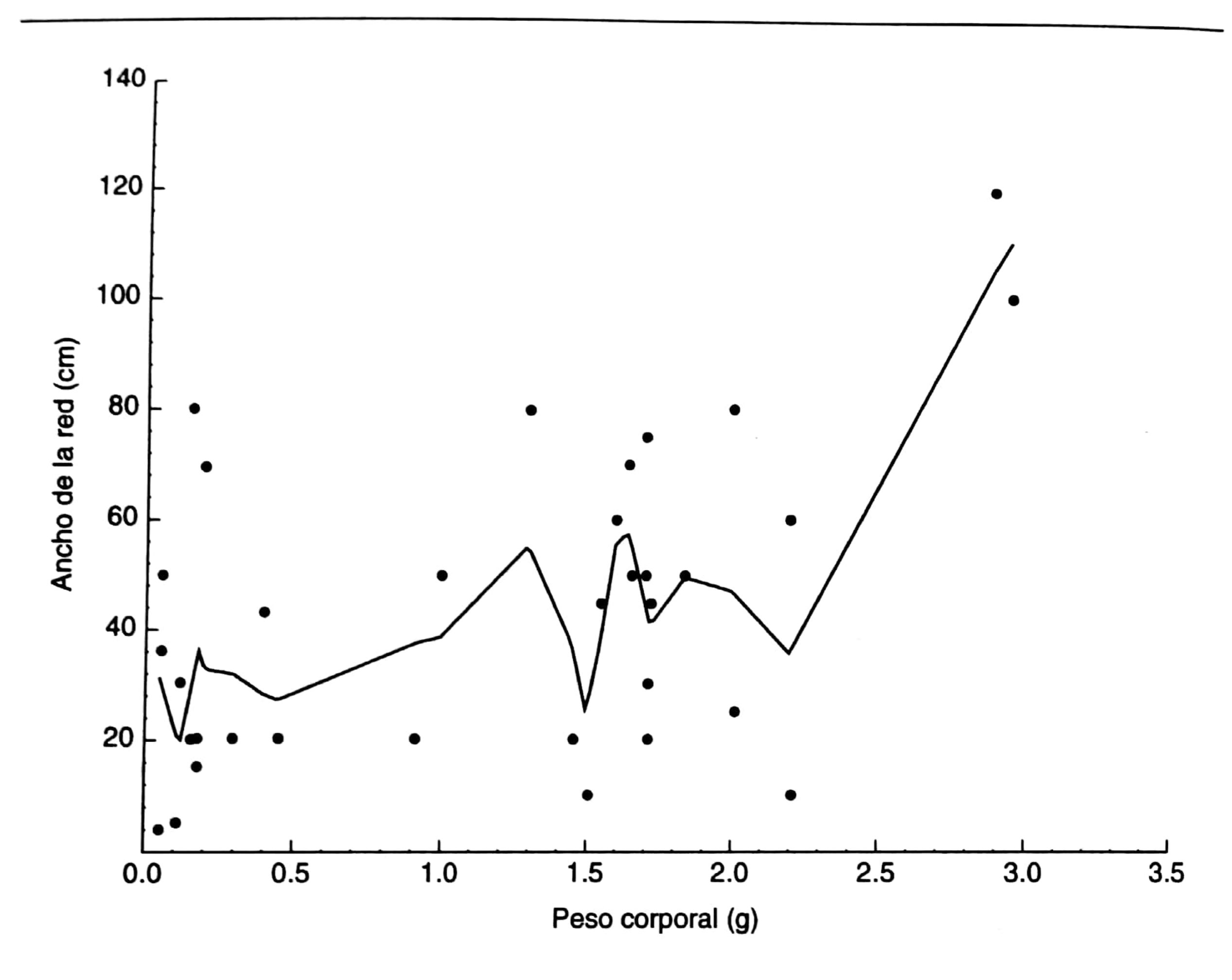


Figura 11.9 Método Lowess aplicado al ejemplo 11.5

Estas herramientas son esencialmente exploratorias o de descripción gráfica de la relación entre las dos variables y se usan cuando no es posible hallar una función de modelación entre ellas. Se aplican, además, a datos diferentes de los de regresión como serían los datos binarios, los de supervivencia y los de estudios de casos y controles.

11.9 Valores óptimos de *X* y número de repeticiones

Cuando la regresión se obtiene de datos experimentales, los niveles o valores de la variable independiente *X* están bajo el control del experimentador. Este deberá hacerse algunas preguntas relacionadas con dicha variable, las cuales le ayudarán en el diseño. Entre otras se tienen:

1. ¿Cuál es el rango de interés para la variable explicativa? O sea, ¿cuáles son los valores extremos de *X*?
2. ¿Cuántos valores de *X* deberán investigarse?
3. ¿Cuántas observaciones deberán tomarse para cada valor de *X*?

No es posible obtener una respuesta única a estas preguntas porque los propósitos de la regresión son distintos y, por consiguiente, cada uno lleva a una respuesta diferente.

Uno de los objetivos principales es el de obtener una relación empírica entre Y y X; por tanto, deben estimarse los coeficientes de la regresión junto con su variabilidad. Se puede observar que los valores de X afectan, por ejemplo, la variabilidad de b; la estimación de su varianza es

$$S_b^2 = \frac{MCE}{\sum (x_i - \bar{x})^2} .$$

Ahora bien, la varianza de la pendiente se minimiza si la suma del denominador se maximiza; esto se logra cuando se toman dos valores de X, uno en cada extremo del rango, y se hace la mitad de las observaciones en cada uno de estos dos puntos.

El rango de X deberá ser bastante amplio para que las inferencias tengan sentido y sean aplicables, pero a la vez, suficientemente estrecho para que el modelo sea lo más simple posible. La decisión es difícil y el investigador deberá consultar otras experiencias de estudios similares. Si el objetivo es estimar una respuesta promedio o predecir un valor de Y para un X_d dado, lo mejor es elegir el rango de X de tal manera que $X_d \approx \bar{X}$.

Si el modelo de respuesta, esto es, el modelo poblacional, se supone lineal, bastan dos valores de X en los extremos del rango. Este diseño tiene la desventaja de que no permite chequear la curvatura en el caso de una respuesta curvilínea.

De acuerdo con Cox (1992), si el objetivo es investigar la existencia de regresión lineal entre las dos variables, bastan dos valores de X. Si se desea una descripción de la curva de respuesta mediante la pendiente y la curvatura, deben usarse tres valores de X. Solo cuando se requiere estimar la forma detallada de la curva de respuesta es necesario mayor número de valores para X.

Con muy pocas excepciones, lo recomendable es la elección de valores de X igualmente espaciados y con igual número de repeticiones (tres o cuatro) en cada uno de estos valores. Las repeticiones permiten probar la bondad del ajuste y su número está supeditado a la importancia del problema y a los costos que involucra su solución. Más concretamente, para probar normalidad de errores y para hacer pruebas de hipótesis sobre algunos parámetros, es aconsejable el tamaño de muestra usual ($n > 30$).

11.10 Correlación lineal simple

Mientras que en el análisis de regresión se investiga la dependencia de una variable de respuesta Y con respecto a otra variable X mediante una ecuación, en la correlación se analiza la interdependencia o covariación entre dos variables mediante un índice o coeficiente. Dicho coeficiente mide la intensidad de la correlación o de la asociación lineal entre ambas variables. Las dos variables deben ser aleatorias y de poblaciones homogéneas y, además, no se clasifican en dependientes e independientes.

Entre los coeficientes de correlación existentes en la literatura estadística, uno de los más conocidos es el coeficiente propuesto por K. Pearson en 1896 y conocido como *coeficiente de correlación producto-momento*. Su expresión matemática para una muestra de n pares de valores de las variables Y_1, Y_2 es:

$$r_{12} = \frac{S_{12}}{\sqrt{(SC_1)(SC_2)}} = \frac{\text{cov}(y_1, y_2)}{DE(y_1)DE(y_2)}.$$

En esta expresión, el numerador se calcula como S_{xy}, haciendo $Y_1 = X$ y $Y_2 = Y$, y las sumas de cuadrados SC_1 y SC_2 como S_{xx} y S_{yy}, respectivamente. Se usa Y_1, Y_2 en vez de X, Y para insistir en que las dos variables son aleatorias.

El coeficiente r_{12} es una estadística muestral que estima el parámetro ρ_{12} de la distribución Normal bivariada, si las variables Y_1, Y_2 tienen dicha distribución. En otro caso, este coeficiente no es más que un índice matemático sin interpretación real adecuada.

El nombre de *producto-momento* se debe a que este coeficiente se ha definido como el cociente entre la covarianza de Y_1, Y_2 —primer momento centrado de la distribución Normal bivariada— y el producto de sus desviaciones estándares. En adelante se escribirá r y ρ en lugar de r_{12} y ρ_{12}.

En el análisis de correlación se mide la intensidad de la asociación lineal entre las variables. Sin embargo, una correlación alta y significativa estadísticamente, no siempre es indicativa de la existencia de causalidad de una de las variables por la otra, pues algunas variables adicionales pueden ser una causa común de las variables estudiadas.

La influencia de ciertas variables latentes hace difícil, muchas veces, la interpretación de correlaciones simples. Se sabe, por ejemplo, que la edad y el peso de los animales están correlacionados, pero además de la edad hay otras variables, como el tamaño, que también influyen en el peso. De igual manera, dos partes del cuerpo pueden estar correlacionadas a causa de su participación simultánea en el tamaño general de un animal.

En la literatura se encuentran con frecuencia valores altos de r para variables que no son aleatorias o que no siguen la distribución Normal bivariada y, aún más, se interpretan como verdaderas asociaciones entre las variables cuando en realidad son correlaciones espurias o ilusorias. Este error es común en proyectos en los que se calculan las correlaciones entre todas las variables imaginables y se publican correlaciones significativas casi sin ningún comentario.

Ejemplo 11.6 Calcular el coeficiente de correlación de Pearson para los datos del ejemplo 11.5.

Sea Y_1 el peso corporal y Y_2 el ancho de la red. Entonces,

$S_{12} = bS_{xx} = 15.37 \times 26.954 = 414.283.$

$SC_1 = S_{xx} = 26.954.$

$SC_2 = S_{yy} = SCT = 27024.8.$

$$r = \frac{414.283}{(26.954 \times 27024.8)^{1/2}} = 0.4854.$$

El parámetro ρ no tiene unidades y está restringido al intervalo [−1, 1]. Cuando su estimador r es positivo y alto, la asociación lineal es directa; cuando es negativo y cercano a −1, la correlación es inversa. Una correlación cercana a cero indica que no existe interdependencia lineal entre las variables, aunque la puede haber curvilínea. Si el diagrama de puntos de Y_1 contra Y_2 (o de Y_2 contra Y_1) muestra una tendencia lineal, se espera una correlación lineal alta entre las dos variables, Al contrario, un diagrama irregular, con un contorno más o menos circular, indica

que la correlación es baja e insignificante y, por tanto, que las variables son estadísticamente independientes.

El coeficiente de correlación está ligado a la regresión, cuando esta es interpretable, mediante la relación $r = bS_1 / S_2$. Aquí S_1 y S_2 son las desviaciones estándares de Y_2 y Y_1, respectivamente, y b es la pendiente de la ecuación de regresión de la gráfica Y_2 contra Y_1.

Por tanto, el coeficiente r es un múltiplo de b y tiene el mismo signo de la pendiente aunque sus interpretaciones son completamente diferentes.

El coeficiente de correlación también puede obtenerse mediante la raíz cuadrada de R^2, así: $r = \pm\sqrt{R^2}$ asignándole a r el mismo signo de la pendiente b.

Ahora bien, el coeficiente r no tiene una interpretación directa y simple: es posible realizar una prueba de las hipótesis H_0: $\rho = 0$ contra H_1: $\rho \neq 0$ ($\rho < 0$, $\rho > 0$) y concluir que la relación es significativa o no significativa estadísticamente. Por otro lado, el cuadrado de r indica en qué porcentaje una de las variables está asociada a la otra.

Las hipótesis anteriores se prueban mediante la estadística:

$$T = r\Big/\sqrt{\left(1 - r^2\right)/(n-2)}. \qquad (11.13)$$

Si $|\ T\ | \geq t_{\alpha/2,\ n\ -\ 2}$, se rechaza H_0. Para el ejemplo anterior, se tiene que $T = 0.4854\Big/\sqrt{(1 - 0.2356)/34} = 3.24$ y $t_{0.025,34} = 2.032$, entonces se rechaza H_0 y se concluye que la correlación entre el peso corporal y el tamaño de la red, de esas arañas, es significativa.

La precisión de la estimación de ρ puede apreciarse por medio del intervalo del 95% de confianza siguiente:

$$Z \pm z_{\alpha/2}\sqrt{1/(n-3)}.$$

donde $Z_{12} = \left(\frac{1}{2}\right)\ln\left(\frac{1+r}{1-r}\right)$, la cual se conoce como transformación de Fisher. Es necesaria la transformación porque la distribución de r es asimétrica. Entonces, $Z_{12} = (1 / 2)$ ln $(1.4854 / 0.5146) = 0.5300$ y el intervalo es:

$$0.5300 \pm 1.96\sqrt{1/33} = (0.1888,\ 0.8712).$$

Los límites en la escala de r se obtienen aplicando la transformación inversa a cada extremo del intervalo (Z_1, Z_2) anterior. Así:

Límite inferior
$$\begin{aligned} r_1 &= \left(e^{2Z_1} - 1\right)\left(e^{2Z_1} + 1\right) \\ &= (e^{0.3776} - 1) / (e^{0.3776} + 1) \\ &= 0.4588 / 2.4588 = 0.1866, \end{aligned}$$

límite superior
$$\begin{aligned} r_2 &= \left(e^{2Z_2} - 1\right)\left(e^{2Z_2} + 1\right) \\ &= (e^{1.7424} - 1) / (e^{1.7424} + 1) \\ &= 4.7110 / 6.7110 = 0.7020. \end{aligned}$$

El intervalo (0.19, 0.70) es bastante amplio para un coeficiente de correlación,

lo que le resta precisión a la estimación; sin embargo, en estudios biológicos, como el del ejemplo, no se esperan resultados con precisiones altas.

En el análisis de correlación deben tenerse en cuenta dos aspectos que pueden alterar los resultados inferenciales. El primero es que el estimador r es sesgado, es decir, subestima a ρ; para muestras pequeñas ($n < 15$) se recomienda una fórmula corregida, Gill (1978, vol. 1). El segundo es que el error estándar de r —el denominador de (11.13)— puede ser una estimación equivocada cuando las variables Y_1 y Y_2 no tienen la distribución Normal bivariada. Además, el tamaño muestral n influye en las estimaciones, que resultan muy variables y poco confiables cuando la muestra es pequeña.

De otro lado, cada área de la ciencia tiene su propio rango para el coeficiente de correlación; por ejemplo, las características orgánicas y heredadas, presentan, frecuentemente, correlaciones inferiores a 0.80 (Snedecor y Cochran, 1984). Estos autores discuten también otros problemas sobre correlación, como la comparación de ρ con un valor $\rho_0 \neq 0$, y la comparación múltiple entre varios coeficientes de correlación.

Existen otros coeficientes de correlación aplicables a variables que no son continuas o que no tienen la distribución Normal bivariada. El más ampliamente usado es el coeficiente de correlación por rangos, propuesto por C. E. Spearman en 1904 y definido como:

$$r_s = 1 - \frac{6}{\left(n^3 - n\right)} \sum_i D_i^2,$$

donde D_i es la diferencia entre los rangos de los valores de Y_1 y Y_2 para la i-ésima pareja de datos muestrales.

Ejemplo 11.7 Calcular el coeficiente de correlación de Spearman para los datos de la tabla 11.12, en la que Y_1 representa el colesterol en la sangre (mg/100 mL) y Y_2 el calcio depositado en las arterias (mg/100 g de tejido). Y_1 y Y_2 fueron medidas en 12 animales (adaptado de Zar, 1999).

Se tiene entonces, que:

$$r_s = 1 - \frac{6 \times 88.00}{1728 - 12} = 0.6923.$$

Si $n \geq 10$ y $r \leq 0.90$, puede usarse la transformación de Fisher para calcular intervalos de confianza para ρ_s (Zar, 1999). El libro de Zar también trae tablas especiales para probar la significancia de r_s.

El análisis de correlación ha tenido sus mayores aplicaciones en estudios no experimentales, sobre todo en aquellos campos en los que las variables son discretas y ordinales. No obstante, su utilidad ha sido demostrada en todas las áreas de la ciencia, en las que se busca describir el comportamiento conjunto de varias variables. La correlación se utiliza en ecología, genética, psicología, sociología, etc., como un paso intermedio en la aplicación de métodos estadísticos más avanzados. Véanse, por ejemplo Baggaley (1965), Maxwell (1967) y Ludwig y Reynolds (1988).

Ejercicios

11.1 Los siguientes valores son los resultados de un estudio sobre la conges-

Tabla 11.12 Datos de Y_1 y Y_2 y sus rangos parar el ejemplo 11.7

Y_1	Rango de Y_1	Y_2	Rango de Y_2	D_i	D_i^2
298	11	59	10.5	0.5	0.25
303	12	52	8	4.0	16.00
233	1.5	42	7	–5.5	30.25
287	9	59	10.5	–1.5	2.25
236	3	24	2	1.0	1.00
245	5	24	2	3.0	9.00
265	7	40	6	1.0	1.00
233	1.5	32	4	–2.5	6.25
286	8	63	12	–4.0	16.00
290	10	57	9	1.0	1.00
264	6	36	5	1.0	1.00
239	4	24	2	2.0	4.00
					$\Sigma D_i^2 = 88.00$

tión del tráfico (número de automóviles por hora) y la concentración de CO (en ppm), efectuado en un sitio de muestreo para determinar la calidad del aire en una ciudad.

N.º de automóviles:	100	110	125	150	175	190	200	225	250
CO (ppm):	8.8	9.0	9.5	10.0	10.5	10.5	10.5	10.6	11.0
N.º de automóviles:	275	300	325	350	375	400			
CO (ppm):	12.1	12.1	12.5	13.0	13.2	14.5			

a. Construir el diagrama de puntos.
b. Calcular *a* y *b* y escribir la ecuación de regresión ajustada.
c. Calcular la tabla ANOVA y dar el valor de *p* para la prueba F.
d. Interpretar el valor de *b* y construir un intervalo del 95% de confianza para β.
e. ¿Cuál es el rango de validez de esta regresión?
f. Calcular los residuos y graficarlos contra *X*. ¿Qué indica esta gráfica en cuanto a los supuestos del modelo?
g. Probar normalidad de los residuos.

11.2 Gómez (1992) midió —in útero— el crecimiento de varios sistemas orgánicos en fetos en la población de Medellín (Colombia), usando la tecnología ultrasónica. Entre las mediciones hechas se tienen las del diámetro biparietal, DBP, y las del diámetro abdominal transverso, DABT, de los fetos desde la semana 13 hasta la 40 de gestación. Entre los objetivos estaba el estudio del comportamiento de la relación DBP / DABT del embarazo. Los siguientes datos son cocientes promedios de varias pacientes en diferentes semanas. Así:

Semana:	13	15	17	19	21	23	25	27	29
DBP/DABT:	1.21	1.19	1.17	1.21	1.18	1.21	1.12	1.13	1.07
Semana:	31	33	35	37	39				
DBP/DABT:	1.03	1.00	1.01	1.00	0.96				

a. Ajustar los datos a un modelo lineal.
b. Interpretar los resultados de la tabla del ANOVA y del coeficiente de determinación.
c. Interpretar los coeficientes de regresión y estimar un intervalo para β.
d. Chequear los supuestos del modelo lineal.

11.3 Una fábrica destila aire líquido para producir oxígeno. El porcentaje de impureza en el oxígeno se considera relacionado linealmente con la cantidad de impurezas, medida en ppm, en el aire. Estos son los datos obtenidos:

Pureza (%)	93.3	92.0	92.4	91.7	94.0	94.6	93.6	
Contaminación (ppm):	1.10	1.45	1.36	1.59	1.08	0.75	1.20	
Pureza %	93.1	93.2	92.9	92.2	91.3	90.1	91.6	91.9
Contaminación (ppm)	0.99	0.83	1.22	1.47	1.81	2.03	1.75	1.68

a. ¿Puede aceptarse la relación lineal entre las variables?
b. ¿Cómo se interpreta el coeficiente *b* en la ecuación ajustada?
c. ¿Cómo se interpreta el coeficiente de determinación en este caso?
d. ¿Cuál es la pureza del oxígeno para una concentración de impurezas de 1.50 ppm?
e. ¿Cuál es la precisión de la estimación anterior? Calcular un intervalo de confianza.
f. Si la pureza del oxígeno debe ser mayor de 95% ¿Cuál es el contenido máximo permisible de contaminación en el aire?

11.4 González (1991) estudió diversos aspectos bioecológicos de los camarones de agua dulce en la región de Urabá (Colombia). Las siguientes mediciones corresponden al largo total, LT, al peso total, PT, y al número de huevos, NH, de hembras *M. acanthurus*.

LT (cm)	PT (g)	NH
3.6	1.5	527
3.7	1.6	398
5.0	1.8	798
5.0	1.5	1265
5.0	1.8	513
5.2	1.8	1000
5.5	2.4	887
5.6	2.6	1838
6.0	2.7	2211
6.2	2.3	2213

LT (cm)	PT (g)	NH
6.5	3.4	2837
6.6	3.0	2715
6.7	3.5	2966
7.0	3.6	2228
7.2	4.0	3961
7.3	4.6	4640
7.8	8.1	6961
8.3	5.0	4379
8.5	6.8	3950
9.4	7.6	6720

a. ¿Existe dependencia lineal, o de otro tipo, entre el largo total y el número de huevos? ¿Cuál es el modelo apropiado?
b. ¿Existe correlación lineal entre largo total y peso total?

11.5 Gómez (1992) también midió el diámetro transverso del tórax en los fetos. Algunos de sus datos (en mm) se presentan a continuación.

Semana	*13*	*17*	*22*	*26*	*32*	*38*
	27	35	45	60	80	96
	19	36	48	57	82	98
	21	32	44	59	71	98
	17	42	50	67	76	89
	18		50	50		92
			49			

a. Construir la tabla del ANOVA incluyendo el análisis de bondad de ajuste para el modelo lineal.
b. Escribir la ecuación de regresión ajustada y calcular el coeficiente R^2.
c. Estimar en forma puntual y por intervalos el diámetro promedio para la semana 28.

11.6 Rojas y otros (1990) investigaron la capacidad de absorción de metales pesados, entre ellos el cadmio, por la *Pistia stratiotis* L. Las plantas se sembraron en estanques con agua contaminada. Parte del experimento consistió en medir el cadmio absorbido por las plantas (en µg Cd/g planta seca) en las primeras 96 horas luego de haber contaminado el agua con 3.0 ppm de cadmio. En cada tiempo se hicieron cuatro repeticiones y los resultados fueron:

Tiempo (horas)	*0*	*24*	*48*	*72*	*96*
	28.3	195.6	340.0	286.7	514.6
	27.6	217.2	300.0	289.9	518.0
	33.4	274.3	430.9	431.8	664.7
	21.4	269.2	425.5	568.8	650.8

a. Hallar la ecuación de regresión ajustada al modelo lineal.
b. Analizar el ajuste del modelo, esto es, la tabla del ANOVA completa.
c. Interpretar los coeficientes de regresión *a* y *b*.
d. Calcular un intervalo de confianza para β e interpretarlo.
e. ¿Cuál sería el riesgo de extrapolar la absorción promedio más allá de las 96 horas?

11.7 En el estudio morfológico de *Micrathyria ocellata,* Ortega (1992) midió el largo del cuerpo, LC, y el largo del fémur de las patas medias, F2, de varios individuos en cada uno de 11 instares. Los valores (en mm) obtenidos fueron:

LC:	1.17	1.40	1.72	2.31	3.05	4.11	5.58	7.68	10.13	13.24	16.85
F2:	0.28	0.39	0.50	0.63	0.81	0.98	1.24	1.57	1.95	2.48	3.32

Usando la fórmula de Huxley, analizar el tipo de crecimiento del fémur respecto al largo del cuerpo.

11.8 Se utilizaron dos métodos: inmunodifusión radial, IDR, y electroinmunodifusión, EID, para medir la concentración (en g/100 mL) de la inmunoglobulina G en el suero sanguíneo de 10 personas donantes. Los resultados se dan a continuación.

Donante:	1	2	3	4	5	6	7	8	9	10
IDR:	1.3	1.5	0.7	0.9	1.0	1.1	0.8	1.8	0.4	1.3
EID:	1.1	1.6	0.5	0.8	0.8	1.0	0.7	1.4	0.4	0.9

a. Hallar la ecuación de regresión lineal entre los dos métodos.
b. Usar la prueba *t* para probar la significancia de la regresión.
c. Si los métodos producen resultados similares, entonces $\alpha = 0$ y $\beta = 1$. Probar diferencia entre métodos y discutir los resultados.

11.9 En una muestra de 15 pacientes con hiperlipoproteinemia primaria de tipo II, se midieron el peso corporal (en kg) y los niveles de triglicéridos (en mg / 100 mL) antes de iniciar un tratamiento (Daniel, 2004).

Peso corporal:	76	97	83	52	70	67	75	78	70	99	75	78	70	77	76
Niveles de triglicéridos:	139	101	57	56	113	42	84	186	164	205	230	186	160	153	139

a. Calcular el coeficiente de correlación de Pearson.
b. Establecer la significancia de la correlación.
c. Estimar la precisión en la estimación mediante un intervalo de confianza.
d. Calcular el coeficiente de correlación de Spearman y compararlo con el de Pearson.

11.10 Vallejo (1989) estudió la relación entre la abundancia y la salinidad en la población de rotíferos de la Ciénaga Grande de Santa Marta (Colombia) durante nueve meses (agosto de 1988 a mayo de 1989). Los siguientes datos corresponden a la estación de muestreo E2 (centro de la ciénaga). La salinidad

se midió mediante porcentajes y la abundancia como el número de individuos de *B. plicatilis* por litro. Los datos obtenidos se muestran a continuación.

Salinidad	*Densidad*
14.7	463.3
14.3	297.5
8.1	94.4
12.1	67.5
6.5	275.6
6.3	663.8
2.3	308.0
1.9	37.5
1.9	76.3
2.4	12.5
2.1	1.9
7.7	2.5
10.0	0.0
15.1	0.0
20.8	0.6
20.1	0.0
21.2	0.0
24.9	0.6
25.1	0.0

a. Construir el diagrama de puntos.
b. Discutir la posibilidad de analizar los datos mediante regresión.
c. Ajustar los datos a un modelo transformable al modelo lineal.
d. Analizar la correlación entre abundancia y salinidad. Usar el coeficiente de Spearman.
e. ¿Qué se puede concluir sobre los supuestos del modelo transformado?

12

Variables categóricas

12.1 Datos categóricos

Una variable categórica es aquella que se mide en una escala cualitativa, o sea, en una escala compuesta de varias categorías (Agresti, 2002). Por ejemplo, el hecho de fumar puede catalogarse en no fumador, ex fumador y fumador; el estado de salud de una persona puede ser sano, recuperado parcialmente, estacionario y enfermo grave. La escala también puede ser ordinal, como la de las clases sociales o la contaminación de un río: no contaminado, contaminación baja y contaminación alta.

Las frecuencias o conteos de las observaciones que se realizan en cada categoría de una variable de respuesta se llaman *datos categóricos*. Los métodos estadísticos para el análisis de estos datos siguen las mismas ideas básicas del diseño experimental; aquí se considerarán, en particular, los dos problemas clásicos de la estadística: comparación y asociación.

El problema de comparación es el de confrontar las distribuciones de frecuencias de las categorías de una variable; por ejemplo, la distribución de los grupos sanguíneos o de las clases sociales en una población bien definida. En general, los procedimientos para comparar distribuciones de variables categóricas son rápidos y sencillos de aplicar, lo cual ha hecho que se extiendan también a variables discretas y continuas que se convierten en categóricas mediante la agrupación en intervalos. El problema de la comparación de k muestras o distribuciones de frecuencias de una variable categórica, se analiza mediante la prueba *ji*-cuadrado de bondad de ajuste.

El problema de asociación estudia la relación entre dos o más variables categóricas definidas en una población. La asociación se describe mediante la distribución conjunta de las dos variables, que resulta de clasificar los individuos de una muestra aleatoria según las categorías de cada variable. El número de individuos o frecuencia en cada clasificación cruzada determina el patrón de asociación de tales variables.

Las categorías de las variables en estudio deben ser exhaustivas y mutuamente excluyentes; es decir, deben definirse suficientes categorías para ubicar todos los individuos de la muestra y cada individuo se asigna a una sola categoría de cada variable.

Para dar un ejemplo, se espera asociación entre el grupo sanguíneo y algunas enfermedades de la infancia: cada infante proporciona información sobre el tipo de sangre y el estado de salud, estos datos se analizan en un arreglo bidimensional o tabla de dos vías. Los valores así obtenidos se llaman *datos de enumeración*. El análisis de asociación o de contingencia entre dos o más variables categóricas, se realiza con la prueba *ji*-cuadrado de independencia.

12.2 Prueba de bondad de ajuste multinomial

En esta sección se verá una prueba de bondad de ajuste introducida por Karl Pearson en el año 1900. Esta prueba, aplicada a datos categóricos, fue revolucionaria en su tiempo y fue, a la vez, uno de los primeros métodos estadísticos inferenciales que se aplicó en estudios científicos. La prueba permite establecer si las probabilidades en una distribución multinomial son equivalentes a valores teóricos supuestos o a valores teóricos conocidos. Se le conoce también como prueba de homogeneidad o prueba *ji*-cuadrado de Pearson.

La hipótesis nula para esta prueba se formula como la igualdad entre los k parámetros $\{p_j\}$ de la distribución y k valores fijos $\{p_{0j}\}$ supuestos por el investigador. Cuando la hipótesis nula es verdadera, las frecuencias esperadas por categoría de la variable son $E_j = np_{0j}$. Si las frecuencias en la muestra, o sea, las frecuencias observadas, son $\{O_j\}$, entonces la estadística propuesta por Pearson para probar H_0 es:

$$X^2 = \sum (0_j - E_j)^2 / E_j \qquad (12.1)$$

Se ha demostrado que para muestras grandes, X^2 tiene una distribución que se aproxima a la *ji*-cuadrado con $k - 1$ *gl*. Si las diferencias $O_j - E_j$ son grandes, el valor de X^2 es igualmente grande para un tamaño muestral n fijo. Valores grandes de X^2 indican falta de ajuste; es decir, existe divergencia entre los valores observados y los teóricos o predichos. La regla de decisión para la prueba es rechazar H_0 si $X^2 \geq \chi^2_{\alpha, k-1}$.

A menudo se obtienen muestras de datos categóricos con el fin de inferir si la población de la cual provienen se ajusta a una distribución teórica o hipotética. Entre las muchas aplicaciones de este proceso, se conocen los estudios genéticos que prueban las teorías de Mendel sobre la herencia natural.

Ejemplo 12.1 En un experimento sobre herencia mendeliana, se cruzaron plantas don diego de la noche de flores rojas y flores blancas, es decir, de especies puras. En la generación F_1 siguiente las flores fueron rosadas, se considera que hubo polinización aleatoria; en la generación F_2 se encontraron 145 flores rojas, 289 rosadas y 138 blancas. ¿Confirman estos resultados la hipótesis de que los colores de las flores deben darse en la proporción 1:2:1?

Las hipótesis estadísticas se plantean de la siguiente manera:

H_0: $P_1 = 1/4$ flores rojas, $P_2 = 2/4$ flores rosadas, $P_3 = 1/4$ flores blancas.

H_1: las probabilidades son diferentes de las anteriores.

El valor de X^2 se calcula en la tabla 12.1 y los valores esperados se obtienen multiplicando cada probabilidad por n, así: $E_1 = 572 \times 1/4 = 143$, etc. Se tienen $k = 3$ categorías o

colores; luego $\upsilon = 3 - 1 = 2$ *gl.* El valor tabulado es $\chi^2_{0.05,2} = 5.99$. Entonces, no se rechaza H_0 y se concluye que los resultados concuerdan con la hipótesis mendeliana ($p = 0.89$).

Un procedimiento alterno a la prueba de bondad de ajuste anterior es la prueba *ji*-cuadrado de la razón de verosimilitud. Su nombre se deriva de la *razón o* cociente entre dos probabilidades que permite concluir qué tan bueno o malo es el ajuste entre una distribución muestral y otra teórica. La estadística de prueba es la expresión G, escrita a veces como G^2, sugerida por S. S. Wilks en 1935 y que se calcula como:

$$G = 2 \sum O_j \ln (O_j / E_j). \qquad (12.2)$$

La estadística G tiene una distribución aproximadamente *ji*-cuadrado con $k - 1$ *gl,* para n moderadamente grande. Para los datos del ejemplo 12.1 este valor es:

$$\begin{aligned} G &= 2[145 \ln(145 / 143) + 289 \ln(289 / 286) + 138 \ln(138 / 143)] \\ &= 2\,[2.014 + 3.016 - 4.911] \\ &= 0.237. \end{aligned}$$

El resultado es muy similar al valor de X^2 y la conclusión es la misma que la anterior: no se rechaza H_0 porque $G < \chi^2_{0.05,2}$.

Tanto X^2 como G tienen distribuciones que convergen a la *ji*-cuadrado con $k - 1$ *gl,* pero no se conoce un valor de n a partir del cual pueda aceptarse una de las pruebas satisfactoriamente; la aproximación depende básicamente de n y de las frecuencias esperadas E_j. Si la muestra es muy pequeña, se recomienda aplicar la distribución multinomial directamente. Algunos autores dicen que si resultan frecuencias esperadas menores que cinco, la aproximación no es buena.

En el caso binomial ($k = 2$), se acostumbra hacer una corrección por continuidad, sugerida por F. Yates, en la forma:

$$X^2 = 2 \sum (|O_j - E_j| - 0.5)^2 / E_j. \qquad (12.3)$$

Otras consideraciones acerca del tamaño de la muestra se estudiarán en la sección 12.5

Ejemplo 12.2 Se colocaron cinco trampas en diferentes lugares de un bosque y se registró, durante tres meses, el número de ratones silvestres capturados en cada una de ellas. El bosque fue dividido inicialmente en 20 cuadros y se eligieron aleatoriamente cinco para colocar las trampas. El objetivo era observar la distribución de los ratones en el bosque, teniendo como hipótesis una distribución uniforme. Los datos obtenidos se muestran en la tabla 12.2 (adaptado de Bishop, 1966).

Tabla 12.1 Valores observados, esperados y de X^2 para los datos del ejemplo 12.1

Color	*Frecuencia observada* O_j	*Frecuencia esperada* E_j	$O_j - E_j$	$(O_j - E_j)^2 / E_j$
Rojas	145	143	2	0.028
Rosadas	289	286	3	0.031
Blancas	138	143	−5	0.175
Total	$n = 572$	572		$X^2 = 0.234$

Tabla 12.2 Datos obtenidos para el ejemplo 12.2

	Trampas					
	1	*2*	*3*	*4*	*5*	*Total*
Número de capturados O_j	23	7	25	19	21	95
Número de esperados E_j	19	19	19	19	19	95

Las hipótesis estadísticas son:
H_0: $p_j = 1/5$, con $j = 1, 2, 3, 4, 5$.
H_1: no todas las probabilidades son iguales entre sí.
Los valores esperados son $E_j = np_j = 95/5 = 19$. Entonces,

$$G = 2[(23 \ln(23/19) + \cdots + 21 \ln(21/19)]$$

$$= 2[4.394 - 6.990 + 6.861 + 0.000 + 2.102] = 12.734, p = 0.012.$$

El valor tabulado ($\chi^2_{0.05,4} = 9.49$) es menor que G, por lo que la hipótesis de distribución uniforme no es la apropiada para estos datos.

Una exploración rápida de los datos permite observar que la segunda trampa capturó un número inferior de ratones, comparado con el resto de las trampas; este resultado puede explicarse averiguando qué ocurrió con esa trampa. Si esta no funcionó correctamente durante todo el tiempo, el dato debe desecharse y el análisis se repite para las cuatro trampas restantes. Bajo esta conjetura: $G = 0.912$, $\chi^2_{0.05,3} = 7.81$, $p = 0.82$, y la hipótesis de investigación puede aceptarse.

El razonamiento anterior es equivalente al rechazo de un dato dudoso en el ANOVA o en la regresión. El dato puede rechazarse si existe una causa comprobada que lo justifique.

La discusión presente sigue las pautas establecidas en el ANOVA de una vía. Si la prueba F era no significativa, terminaba el análisis; en caso contrario, se continuaba con la comparación de medias o varianzas de los tratamientos. Aquí ocurre lo mismo con la prueba *ji*-cuadrado; si esta es significativa, se procede con un análisis detallado de la tabla de frecuencias observadas y esperadas. De acuerdo con el primer resultado, el paso siguiente es la comparación por pares de las frecuencias de las diferentes categorías.

Ejemplo 12.3 En el ejemplo anterior se concluyó que había diferencias significativas entre las proporciones p_j, de ratones capturados por trampa.

La comparación de estas proporciones sigue un procedimiento similar al de Tukey (sección 6.8), pero usando las transformaciones angulares de las proporciones muestrales (véase tabla 12.3). La fórmula para la transformación (Zar, 1999) es:

$$p_i^* = (1/2)\left[\text{arcsen}\sqrt{x/(n+1)} + \text{arcsen}\sqrt{(x+1)/(n+1)}\right].$$

Por ejemplo,

$$p_1^* = (1/2)\left(\text{arcsen}\sqrt{25/96} + \text{arcsen}\sqrt{26/96}\right) = 31.01,$$

$$p_2^* = (1/2)\left(\text{arcsen}\sqrt{23/96} + \text{arcsen}\sqrt{24/96}\right) = 26.97,$$

y así sucesivamente.

Tabla 12.3 Proporciones transformadas según el ejemplo 12.3

Trampas	*3*	*1*	*5*	*4*	*2*
Proporciones ordenadas	25/95	23/95	21/95	19/95	7/95
Proporciones transformadas	31.0	29.7	28.3	26.8	16.2

La estadística de comparación es $W = q_{\alpha,\,\infty,\,k}\,EE$. El valor tabulado se lee en la tabla A.12 y es ($q_{0.05,\,\infty,\,5} = 3.86$). El error estándar de la diferencia se calcula como:

$$EE = \sqrt{820.70/(n+0.5)} = \sqrt{820.70/(95+0.5)} = 2.93.$$

Cuando las proporciones tienen tamaños de muestra n_i y n_j, se usa

$$EE = \left[410.35/(n_i+0.5) + 410.35/\left(n_j+0.5\right)\right]^{1/2}.$$

Las diferencias entre proporciones transformadas se comparan con $W = 3.86 \times 2.93 = 11.31$. En la tabla 12.4 se muestran los resultados y se concluye que la trampa 2 presentó un resultado similar al de la 4, pero inferior con respecto a las demás trampas.

12.3 Tablas de contingencia r × c

La prueba *ji*-cuadrado de homogeneidad, tratada en el numeral anterior, compara k atributos de una variable y las frecuencias absolutas se presentan en una tabla de la forma $1 \times k$, es decir, una fila y k columnas; el objetivo es la comparación de varias probabilidades.

El concepto de la asociación involucra la relación entre dos variables categóricas en la misma población. Por lo cual la información necesaria para el análisis se presenta en una tabla de dos vías que contiene las frecuencias absolutas observadas.

Sean A y B dos variables de respuesta categóricas, la variable A con r atributos o niveles y la variable B con c atributos. Cuando se clasifican los individuos sobre ambas variables, se tienen rc combinaciones posibles. Las respuestas de los individuos, elegidos aleatoriamente de una población, tienen una distribución de probabilidades representada en una tabla rectangular con r filas y c columnas, formando rc celdas. Cuando en las celdas se tienen frecuencias observadas, se habla de una tabla de contingencia o de clasificación cruzada, tabla 12.5.

Se acostumbra representar por n_{ij} la frecuencia observada en la celda (i,j) y por p_{ij} $(= n_{ij}/n)$ la probabilidad de que una observación caiga en la celda (i,j). Además,

Tabla 12.4 Tabla de comparaciones de proporciones

Trampas	*Diferencia*	*Conclusión*
2 contra 3	14.8	Diferencia significativa
2 contra 1	13.5	Diferencia significativa
2 contra 5	12.1	Diferencia significativa
2 contra 4	10.6	Diferencia no significativa
4 contra 3	4.2	Diferencia no significativa

se supone que las frecuencias n_{ij} tienen una distribución multinomial con probabilidades conjuntas p_{ij} para las variables A y B.

Las distribuciones marginales son los totales de filas y columnas obtenidos sumando las probabilidades conjuntas. La tabla 12.5 representa, esquemáticamente, una tabla de contingencia de dos vías. En esta tabla $n = \sum\sum n_{ij}$ además, $p_{i.}$ y $p_{.j}$, son las probabilidades marginales de filas y columnas, respectivamente.

Se dice que las variables A y B son estadísticamente independientes cuando las probabilidades conjuntas son iguales a los respectivos productos de sus probabilidades marginales, esto es, si:

$p_{ij} = (p_{i.})(p_{.j})$ para todo par (i,j), con $p_{i.} = n_{i.} / n$ y $p_{.j} = n_{.j} / n$.

Para probar asociación (o lo contrario: independencia) en tablas de contingencia de dos vías, también se utiliza la estadística X^2 de Pearson, pues tanto la independencia como la homogeneidad están planteadas en términos de comparaciones de probabilidades. Las hipótesis estadísticas:

H_0: Las variables A y B son independientes.

H_1: Las variables A y B son dependientes o están asociadas

se prueban mediante la estadística:

$$X^2 = \sum_i \sum_j \left(O_{ij} - E_{ij}\right)^2 / E_{ij}, \qquad (12.4)$$

que tiene una distribución aproximadamente *ji*-cuadrado con $\upsilon = (r - 1)\,(c - 1)\,gl$.

Las frecuencias O_{ij} son los valores observados n_{ij} en cada celda de la tabla. Como las frecuencias esperadas son desconocidas, estas se estiman usando los totales marginales, así:

$E_{ij} = (1/n)(n_{i.})(n_{.j})$.

Entonces, se rechaza H_0 si $X^2 \geq \chi^2_{\alpha,\,\upsilon}$

Ejemplo 12.4 Se llevó a cabo un estudio para determinar si el grupo sanguíneo, GS, es

Tabla 12.5 Tabla de contingencia de dos vías o tabla $r \times c$

Variable A	*Variable B*						*Totales*
	1	*2*	...	*J*	...	*c*	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
2	n_{21}						
⋮	⋮						⋮
i	n_{i1}	...		n_{ij}	...	n_{ic}	$n_{i.}$
⋮	⋮						⋮
r	n_{r1}	...		n_{rj}	...	n_{rc}	$n_{r.}$
Totales	$n_{.1}$	...		$n_{.j}$	...	$n_{.c}$	n

un factor de riesgo en el síndrome de muerte súbita infantil, SMSI. En la tabla 12.6 se presentan los datos obtenidos de una muestra de 135 infantes con SMSI y 135 controles (adaptado de Koopmans, 1987).

Las frecuencias esperadas son $E_{11} = (135 \times 124)/270 = 62, \cdots, E_{42} = (35 \times 11)/270 = 5.5$. Entonces:

$$X^2 = \frac{(69-62)^2}{62} + \frac{(40-55.5)^2}{55.5} + \cdots + \frac{(2-5.5)^2}{5.5}.$$

$$X^2 = 0.790 + 4.329 + 2.227 + 0.790 + 4.329 + 2.083 + 2.227$$
$$= 18.858.$$

Los gl de X^2 son $(r-1)(c-1) = 3 \times 1 = 3$ y $\chi^2_{0.05,3} = 7.81$.

Para este ejemplo, el nivel de significación de la prueba es $p = 0.0003$, así que se tienen razones para suponer que el grupo sanguíneo explica, en parte, el SMSI. La prueba *ji*-cuadrado indica que no hay independencia entre las variables, pero no proporciona información sobre la parte de la tabla donde ocurre la falta de independencia.

Una mirada a la tabla de porcentajes (tabla 12.7), calculados con respecto a los tamaños muestrales, permite reconocer el tipo A como el de mayor disparidad (diferencia entre 29.6 y 52.6%). En realidad, su contribución a la suma X^2 es 4.329, que es el mayor de todos los sumandos.

Para saber si, en efecto, el tipo A es el responsable de la significación de la prueba inicial, se debe subdividir la tabla 4×2 en tablas menores, de modo que los correspondientes valores de la suma X^2 particionada sean independientes.

La partición de la estadística X^2 puede hacerse de varias formas según la propiedad aditiva de la distribución χ^2 que puede enunciarse así: una tabla de contingencia que tenga un valor X^2 con υ gl, puede descomponerse en m tablas con valores X_i^2 y υ_i, gl $(i = 1,2, \ldots, m)$, independientes entre sí, de tal manera que:

$$\sum_i X_i^2 = X_0^2 \text{ y } \sum_i \upsilon_i = \upsilon.$$

Tabla 12.6 Datos del ejemplo 12.4. Las frecuencias esperadas están entre paréntesis

Grupo sanguíneo	*SMSI*		*Controles*		*Totales*
O	69	(62)	55	(62)	124
A	40	(55.5)	71	(55.5)	111
B	17	(12)	7	(12)	24
AB	9	(5.5)	2	(5.5)	11
Totales	135		135		270

Tabla 12.7 Tabla de porcentajes para el ejemplo 12.4

GS	*SMSI*		*Controles*	
O	51.1	%	40.7	%
A	29.6		52.6	
B	12.6		5.2	
AB	6.7		1.5	

De acuerdo con la disparidad en los porcentajes del grupo sanguíneo A, la tabla 12.6 se puede subdividir de la siguiente manera:

GS	*SMSI*	*Controles*
O	69	55
A	40	71
B	17	7
AB	9	2

$X^2 = 18.858$

$\upsilon = 3$

$p = 0.00030$

⇒

GS	*SMSI*	*Controles*
O	69	55
B	17	7
AB	9	2

$X_1^2 = 4.326$

$\upsilon_1 = 2$

$p = 0.115$

+

GS	*SMSI*	*Controles*
O + B + AB	95	64
A	40	71

$X_2^2 = 14.701$

$\upsilon_2 = 1$

$p = 0.00013$

No hay asociación de grupos sanguíneos O, B y AB con enfermedad ($X^2 = 4.326$, $p = 0.115$). Se confirma, entonces, que la asociación se debe a que la frecuencia de infantes del tipo A es menor en el grupo SMSI que en los controles.

El resultado anterior se interpreta diciendo que los recién nacidos con sangre tipo *A* tienen menos riesgo de morir de SMSI que el resto de los infantes. Infortunadamente, con este estudio nada puede decirse sobre el grupo sanguíneo de los infantes con mayor riesgo de muerte por el SMSI.

Este método —hacer comparaciones en subtablas— es útil en trabajos exploratorios, sobre todo cuando el investigador encuentra una X^2 significativa y no tiene ideas claras sobre la dirección en que se desvía la hipótesis inicial. Las conclusiones obtenidas del análisis entre subtablas son provisionales, ya que fueron elaboradas después de consultar los datos. Por tanto, las hipótesis nuevas deberán contrastarse con datos obtenidos de muestras posteriores.

Cuando se realizan *m* pruebas separadas en subtablas de una tabla de contingencia mayor, en la que se ha rechazado la hipótesis de independencia, se recomienda seguir el método de Bonferroni para fijar el error α, ya que las subtablas pueden ser redundantes. Wickens (1989) trae tablas para la prueba de Bonferroni.

Por otra parte, es necesario aclarar que la suma de los X_i^2 no es exactamente igual a X^2 pues, aunque no hubiera redondeo de cifras, podría haber una pequeña diferencia entre estos totales. Agresti (2002) sugiere el uso de la estadística *G* de Wilks en vez de X^2, porque los valores de *G* para las subtablas suman exactamente la *G* de la tabla original.

La partición de una tabla de tamaño *rc* en tablas menores no puede hacerse de cualquier manera: se requiere que las componentes X_i^2 sean independientes y tengan la distribución *ji*-cuadrado. Kimball (1954) ha sugerido algunas fórmulas para la partición de X^2, mientras que Snee (1974) ha descrito algunos métodos gráficos sobre la subdivisión de tablas de contingencia. Ejemplos de particiones de tablas de contingencia se encuentran en varios textos, como los de Wickens (1989), Steel y Torrie (1997) y Everitt (1992).

12.4 Tablas de contingencia 2 × 2

Las tablas 2 × 2 representan el caso más simple para investigar asociación entre dos variables. Su estructura puede generalizarse como en la tabla 12.8 y la fórmula para la estadística de prueba dada por (12.4) se reduce a la expresión siguiente:

$$X^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(c + d)(a + b)}. \qquad (12.5)$$

Ejemplo 12.5 En un estudio sobre aislamiento de dermatofitos en caninos, realizado durante 1988 en varias clínicas de Medellín, G. E. García y A. I. Gutiérrez encontraron la información que se muestra en la tabla 12.9 (datos cedidos gentilmente por las investigadoras).

Para analizar la asociación entre el sexo y la enfermedad, se calcula primero la expresión (12.5), así:

$$X^2 = \frac{205(26 \times 70 - 11 \times 98)^2}{37 \times 168 \times 81 \times 124} = 1.808.$$

Este resultado se compara con el valor tabulado para $\chi^2_{0.05,1}$ que es 3.84 (p = 0.18). Entonces se acepta la hipótesis nula o de independencia lo que quiere decir que la presencia de dermatofitos no está asociada al sexo de los caninos.

Cuando el tamaño muestral es pequeño, F. Yates sugirió la aplicación de una corrección por continuidad —porque se está usando la distribución *ji*-cuadrado que es continua para analizar datos de enumeración— en la siguiente forma:

$$X^2 = \frac{n(|ad - bc| - 0.5n)^2}{(a + c)(b + d)(c + d)(a + b)}. \qquad (12.6)$$

Ejemplo 12.6 Se cree que el factor Rh está asociado a la artritis en las personas. Una muestra aleatoria de 48 individuos quedó repartida como se muestra en la tabla 12.10.

La estadística usual es:

$$X^2 = \frac{48(11 \times 21 - 6 \times 10)^2}{17 \times 31 \times 27 \times 21} = 4.70,$$

y su nivel de significación es p = 0.030.

La prueba de asociación resulta significativa, indicando que la artritis está asociada al factor Rh. Como la muestra es pequeña puede aplicarse la fórmula corregida, con la cual se obtiene:

$$X_c^2 = \frac{48(|11 \times 21 - 6 \times 10| - 24)^2}{17 \times 31 \times 27 \times 21} = 3.47,$$

y la prueba es no significativa (p = 0.062).

La corrección de Yates produce un valor menor para X^2 y, en este caso particular, dicho valor no alcanza la significancia. Grizzle (1967) mostró que la aplicación de la corrección de Yates resulta casi siempre en una prueba demasiado conservadora, y es innecesaria aún con tamaños de muestra tan pequeños como n = 20. Otros autores como Everitt (1992) la recomiendan para muestras pequeñas, ya que produce resultados similares a los de la prueba exacta de Fisher.

Tabla 12.8 Tabla de contingencia 2 × 2

	A	*No A*	*Totales*
B	*a*	*b*	*a + b*
No B	*c*	*d*	*c + d*
Totales	*a + c*	*b + d*	*a + b + c + d = n*

Tabla 12.9 Datos obtenidos para el ejemplo 12.5

	Positivos	*Negativos*	*Total*
Machos	26	98	124
Hembras	11	70	81
Total	37	168	205

Tabla 12.10 Datos para el ejemplo 12.6

	Con artritis	*Sin artritis*	*Total*
Rh+	11	10	21
Rh–	6	21	27
Total	17	31	48

La prueba *ji*-cuadrado de asociación en tablas 2 × 2 es equivalente a la prueba de la diferencia de dos proporciones. En el ejemplo anterior, el objetivo es comparar la proporción de individuos con artritis y que son Rh+ (11/21 = 0.524) con la proporción de individuos con artritis y que son Rh– (6/27 = 0.222). Si la artritis es independiente del factor Rh, se espera que la proporción de personas Rh+ que tienen artritis sea similar a la proporción de personas Rh– que también tengan la enfermedad. Si estas proporciones son diferentes, la artritis puede considerarse asociada a la presencia del antígeno (la mayor de las dos proporciones).

Una estimación del error estándar de la diferencia de estas proporciones es 0.139 y un intervalo aproximado del 95% de confianza para la diferencia es 0.302 ± 1.96 × 0.139 = (0.029, 0.574). Como el intervalo no contiene el cero se concluye que las proporciones son diferentes. Este resultado confirma el ya obtenido con la prueba *ji*-cuadrado sin la corrección de Yates. Ahora bien, mediante *software* estadístico se obtuvo $X^2 = 4.58$ y $p = 0.038$ corroborando la significancia de las pruebas anteriores. El computador realiza una prueba exacta usando una distribución hipergeométrica.

12.5 Tamaño de la muestra

En este punto se debe plantear el problema del tamaño muestral en tablas de contingencia. Como ya se mencionó, la estadística X^2 o la G, convergen en distribución a la *ji*-cuadrado cuando $n \to \infty$ y para un número N fijo de celdas; es decir, la aproximación depende de n y no de N. Además, se ha demostrado que X^2 es superior a G cuando los tamaños de muestra son pequeños, generalmente cuando $n / N < 5$.

Si las frecuencias esperadas son muy pequeñas, el valor de X^2 es sesgado y sobrestima el valor de la distribución χ^2, entonces se tiende a rechazar H_0 con una probabilidad mayor que α, situación claramente indeseable. Cochran (1954) estudió la aproximación de X^2 a la distribución *ji*-cuadrado, y concluyó que esta puede aplicarse en tablas con un número de *gl* mayor que 1 cuando ninguna frecuencia esperada sea menor que 1.0 y, a la vez, cuando no más del 20% de las frecuencias esperadas sean menores que 5.0. Esta sugerencia es igualmente válida para la prueba de homogeneidad; sin embargo, para $k = 2$ con E_j pequeña, debe preferirse la prueba binomial.

En tablas 2 × 2, Cochran recomienda la prueba exacta de Fisher cuando $n < 20$, o cuando $n < 40$ y una de las frecuencias esperadas es menor que 5.0. En otros casos, la prueba *ji*-cuadrado corregida proporciona resultados satisfactorios.

Los procedimientos estadísticos anteriores son válidos en la medida en que se satisfagan los supuestos del muestreo. En particular, para datos categóricos se requiere que el muestreo sea multinomial, esto es, que las n observaciones se tomen independientemente y de la misma población en la cual se han definido las N categorías. A veces una de las variables puede considerarse variable explicativa, como el sexo para el estudio de algunas enfermedades. En estos casos, no solo la n es fija sino también los totales marginales de la variable explicativa, por ejemplo, 100 machos y 100 hembras.

En estudios prospectivos, los totales de filas, o sea la variable explicativa, son generalmente fijos; en estudios retrospectivos son fijos los totales de columnas, esto es, las variables de respuesta; en estudios transversales solamente el total n de la muestra es fijo. Para estos estudios se aplica la misma prueba *ji*-cuadrado, pero la interpretación depende de las hipótesis planteadas en el diseño inicial.

Una solución al problema de frecuencias esperadas pequeñas cuando la tabla es grande, consiste en combinar algunas categorías adyacentes de modo que los totales marginales aumenten. Sin embargo, este método hace menos sensible la prueba y reduce el alcance de la inferencia. Aprovechando las enormes posibilidades para hacer cálculos numéricos que tienen los computadores, los programadores de *software* estadístico han resuelto el problema del tamaño muestral en tablas de contingencia. Programas como el StatXact realizan pruebas exactas de asociación para tablas r × c de tamaños relativamente grandes, usando la prueba de Fisher-Freeman-Halton, Mehta y Patel (1986). Para tablas muy grandes se utilizan técnicas de simulación del tipo Monte Carlo que producen resultados muy cercanos a los de las pruebas exactas.

12.6 Prueba exacta de Fisher

Bajo la hipótesis de independencia, la distribución exacta de las frecuencias en una tabla 2 × 2 es la hipergeométrica, la cual resulta de condicionar o fijar las frecuencias marginales en ambos márgenes. La probabilidad exacta de que en una tabla 2 × 2 con totales marginales fijos ocurran las frecuencias a, b, c, d observadas, se calcula, Fisher (1970), como:

$$p = \frac{(a+c)!(b+d)!(c+d)!(a+b)!}{a!b!c!d!n!}.$$

El signo ! indica el factorial del número, por ejemplo, 4! = 4 × 3 × 2 × 1 = 24 y 0! = 1.

La prueba *ji*-cuadrado de independencia es equivalente a la prueba bilateral de igualdad de dos proporciones, donde la hipótesis alternativa es de la forma H_1: $p_1 \neq p_2$. La prueba exacta de Fisher, en cambio, puede ser bilateral, como la anterior, o unilateral.

Ejemplo 12.7 Suponiendo que la muerte por causas cardiovasculares, CV, está asociada a una dieta alta en sal, se realizó un estudio retrospectivo que arrojó los datos de la tabla 12.11.

Se quiere evaluar si la probabilidad de muerte por causas cardiovasculares es mayor para las personas con dieta alta en sal que con dieta baja en sal.

1. Prueba unilateral. La prueba unilateral tiene como hipótesis alternativa $H_1: p_1 > p_2$ o $p_1 < p_2$. El primer paso en la aplicación de la prueba de Fisher es constatar que las probabilidades obtenidas están en la dirección de la hipótesis H_1. De acuerdo con la tabla 12.11, se tiene que $p_1 = 5/9$ es mayor que $p_2 = 2/8$, o sea que los datos observados están sugiriendo que la probabilidad de muerte por causas cardiovasculares es mayor entre las personas con dieta

Tabla 12.11 Datos para el ejemplo 12.7

Dieta	*CV*	*No CV*	*Totales*
Alta en sal	5	4	9
Baja en sal	2	6	8
Totales	7	10	17

alta en sal. Si las probabilidades se dan en la dirección contraria a H_1, no tiene sentido continuar la prueba.

Cuando los resultados están en la dirección predicha, la prueba exacta de Fisher responde a la siguiente pregunta:

¿Cuál es la probabilidad de obtener una tabla 2 × 2 como la observada, u otra más extrema en la dirección predicha, cuando la hipótesis nula es verdadera?

De otro lado, para probar una hipótesis unilateral se debe calcular la probabilidad de la tabla correspondiente a la obtenida de la muestra y las probabilidades de todas las demás tablas que representen un resultado más extremo que el observado, en la dirección esperada. La suma de estas probabilidades se compara con el valor de α: si la suma es menor o igual que este valor, se rechaza H_0.

Para obtener una tabla más extrema que la observada, se le resta 1 a la frecuencia más pequeña de la tabla y se ajustan las otras frecuencias de modo que los totales marginales no varíen; el proceso se repite hasta obtener una celda con frecuencia cero, la cual será la tabla más extrema de todas. Para el ejemplo en discusión, existen dos tablas más extremas que la observada, las tablas 12.12 (b) y (c).

En la tabla observada se tiene un riesgo de cinco en nueve para el grupo de dieta alta en sal. Los resultados más extremos en la misma dirección, o sea, en favor de dieta alta en sal, son de seis en nueve, tabla (b) y de siete en nueve, tabla (c).

La probabilidad de obtener cada una de estas tablas es:

$$p_a = \frac{7!10!8!9}{5!4!2!6!17!} = 0.1814.$$

$$p_b = \frac{7!10!8!9}{6!3!1!7!17!} = 0.0346.$$

$$p_c = \frac{7!10!8!9}{7!2!0!8!17!} = 0.0018.$$

La probabilidad de obtener una tabla como la observada o una más extrema, es:

$$p = p_a + p_b + p_c = 0.1814 + 0.0346 + 0.0018 = 0.2178.$$

Esta probabilidad es mayor que 0.05 y, por tanto, no se rechaza H_0. Puede aceptarse que la probabilidad de morir por causas cardiovasculares es similar para ambas dietas.

2. *Prueba bilateral.* La prueba bilateral tiene como alternativa que las probabilidades son diferentes. Para el ejemplo anterior, se formula la hipótesis de que la probabilidad de muerte para los de dieta alta en sal es diferente de la probabilidad de muerte para los de dieta baja en sal. Claro que la hipótesis formulada en esta forma no tiene mucho sentido, pues se cree que la dieta rica en sal es perjudicial y no que sea eventualmente benigna para el sistema cardiovascular.

Para ilustrar la prueba bilateral, se acepta que tiene sentido la hipótesis alternativa $H_1: p_1 \neq p_2$. Para probar esta hipótesis se requiere calcular la probabilidad exacta de cada una de las tablas más extremas, pero en la otra dirección: se suman las probabilidades obtenidas en ambas direcciones y la suma se compara con α. Si las filas o las columnas

Tabla 12.12 Tablas extremas para el ejemplo 12.7

Dieta	*CV*	*No CV*	*CV*	*No CV*	*CV*	*No CV*
Alta en sal	5	4	6	3	7	2
Baja en sal	2	6	1	7	0	8
	(a)		(b)		(c)	

son iguales entre sí, la probabilidad bilateral es dos veces la probabilidad unilateral. Si los totales marginales son diferentes, y dado que a veces es difícil reconocer las tablas más extremas en la otra dirección, entonces se compara la suma de probabilidades en una sola dirección con $\alpha/2$. La prueba resultante es aproximada ya que las probabilidades no son simétricas en las dos direcciones.

Según Zar (1999), para hallar las tablas más extremas en la otra dirección, se toma la tabla más extrema en la dirección ya analizada; esto es, la tabla que contiene una frecuencia igual a cero. Luego se localiza la celda correspondiente a los totales mínimos de fila y columna. La frecuencia de esta celda se resta del total marginal menor y con este resultado se construye una nueva tabla. Esta es la tabla más extrema en la dirección contraria. Si la probabilidad para esta tabla es menor que la probabilidad de la tabla observada, entonces se disminuye en 1 la frecuencia de esta celda y se construye otra tabla menos extrema. Se calcula la probabilidad de esta nueva tabla y se repite el proceso hasta que la probabilidad de la nueva tabla sea mayor que la de la tabla originalmente observada. El proceso es sencillo, como lo muestra el ejemplo a continuación.

Ejemplo 12.8 Regresando al ejemplo 12.7, allí se encontró que la tabla más extrema era la tabla 12.12 (c) que se reproduce como la tabla 12.13.

El total menor de fila es 8 y de columna es 7, y la frecuencia de celda correspondiente es 0. En esta celda se coloca la diferencia 7 – 0 = 7, ya que siete es el total marginal menor.

La tabla más extrema en la otra dirección es la tabla 12.14 (a) y su probabilidad es $p_a = 0.00041 < 0.1814$. Se disminuye en 1 la frecuencia de la celda afectada, o sea se cambia el siete por seis para formar una tabla nueva, tabla (b). Se continúa el proceso obteniéndose $p_{b'} = 0.01296$ y $p_c = 0.10366$. Como $p_d = 0.30237 > 0.1814$, esta última tabla no es de las extremas y no se tiene en cuenta. Entonces,

Probabilidad en la otra dirección $= p_a + p_b + p_c = 0.11703$.

Probabilidad bilateral exacta = 0.2178 + 0,1170 = 0.3348. En la figura 12.1 se gráfica la distribución de probabilidades completa para el ejemplo 12.7.

12.7 Homogeneidad de experimentos repetidos

Un problema de alguna ocurrencia en investigación es el de integrar datos de diferentes experimentos para obtener un solo resultado global. Por ejemplo, cuando cada uno de varios experimentos genera una tabla 2 × 2. La primera idea sería sumar las frecuencias respectivas de cada tabla y formar una tabla 2 × 2 resultante de la combinación de todas ellas, o sea de una supermuestra. Es válida la prueba *ji*-cuadrado para esta tabla, pero el resultado podría no reflejar las características de los experimentos individuales.

Otra manera de analizar los resultados consiste en calcular el valor de X_i^2 para cada experimento y luego sumar estos valores aprovechando la propiedad regenerativa de la distribución *ji*-cuadrado. Se obtiene un valor X^2 total que es la acumulación de las informaciones parciales proporcionadas por cada uno de los experimentos independientes. Si al valor X^2 total se le resta el valor X^2 combinado de la supermuestra, se obtiene una estadística llamada

Tabla 12.13 Tabla más extrema para el ejemplo 12.8

7	2	9
0	8	8
7	10	17

Tabla 12.14 Tablas extremas, según la prueba bilateral, para el ejemplo 12.8

0	9	1	8	2	7	3	6
7	1	6	2	5	3	4	4
(a)		(b)		(c)		(d)	

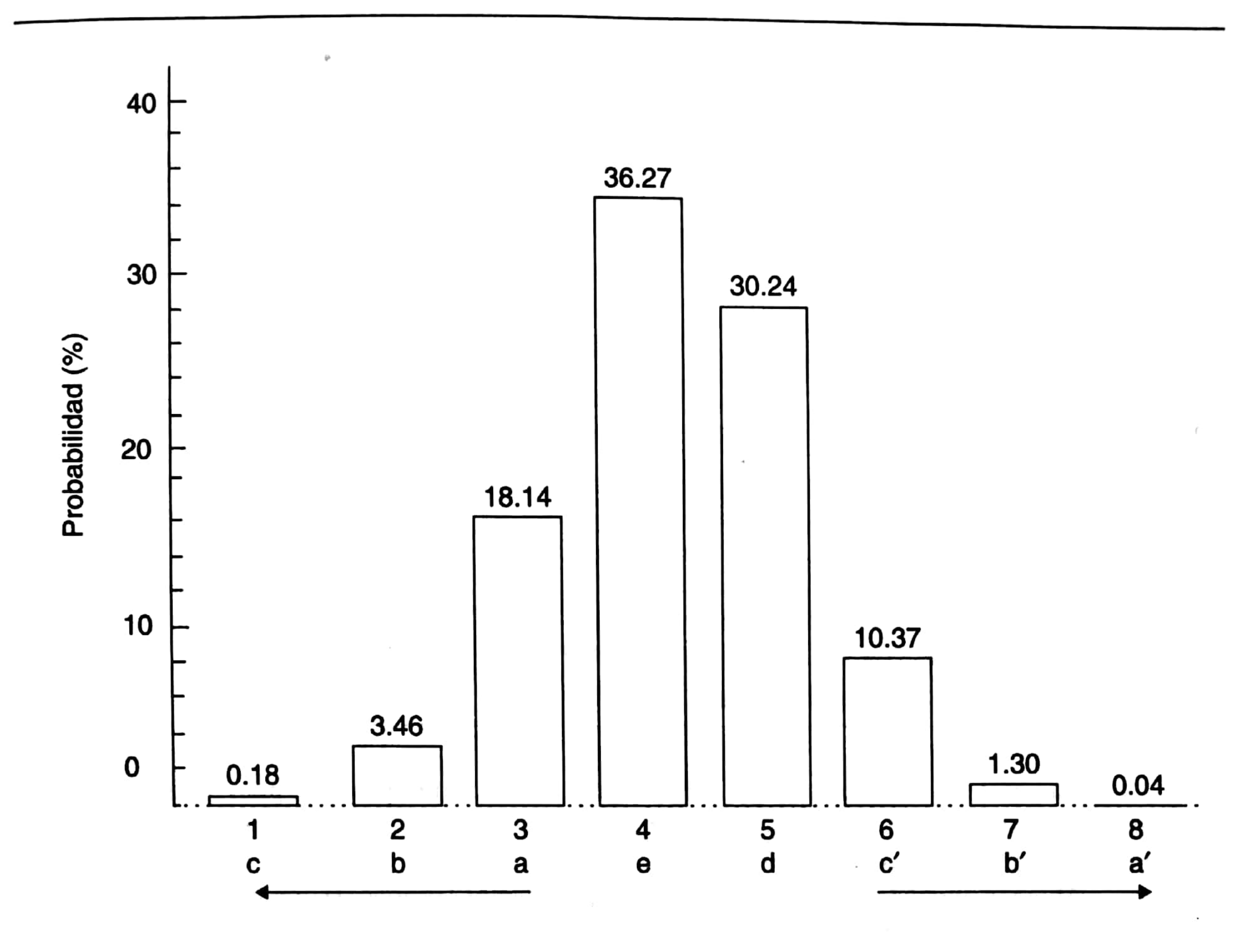


Figura 12.1 Distribución de probabilidades para las tablas posibles del ejemplo 12.6

ji-cuadrado de heterogeneidad, que permite medir la falta de homogeneidad en los resultados de los diferentes experimentos.

Para concretar el razonamiento anterior, se analizará una serie clásica de diez experimentos realizados por Gregor Mendel en 1866.

Ejemplo 12.9 En una serie de experimentos para analizar la forma de las semillas, Mendel observó las frecuencias de semillas redondas y anguladas en diez plantas de la misma especie. Los datos se muestran en la tabla 12.15.

De acuerdo con la teoría genética, las semillas deberán producirse en la proporción de tres redondas por una angulada. Para probar la homogeneidad de la razón 3:1 en los experimentos, se procede a calcular los valores de X^2 para cada uno y luego para los experimentos combinados. Los resultados se resumen en la tabla 12.16.

Cada una de las muestras se probó contra la hipótesis: las semillas redondas y anguladas están en la proporción 3:1. Aquí no se aplica la corrección de Yates.

Tabla 12.15 Datos experimentales para el ejemplo 12.9

Semillas	*Experimentos*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Redondas	45	27	24	19	32	26	88	22	28	25
Anguladas	12	8	7	10	11	6	24	10	6	7

Tabla 12.16 Valores de X^2 para cada experimento y para la muestra combinada del ejemplo 12.9

Experimento	*Semillas redondas*	*Semillas anguladas*	*Total*	X^2	*gl*
1	45	12	57	0.4737	1
2	27	8	35	0.0857	1
3	24	7	31	0.0968	1
4	19	10	29	1.3908	1
5	32	11	43	0.0078	1
6	26	6	32	0.6667	1
7	88	24	112	0.7619	1
8	22	10	32	0.6667	1
9	28	6	34	0.9804	1
10	25	7	32	0.1667	1
Suma de X^2				5.2972	10
X^2 combinado	336	101	437	0.8307	1
	Diferencia (heterogeneidad)			4.4665	9

La hipótesis nula no fue rechazada en ninguna de las diez pruebas, pues ningún X^2 resultó mayor que $3.841 = \chi^2_{0.05,1}$. Por tanto, la conclusión es que dicha proporción es constante en las plantas de esta especie.

Ahora se suman los resultados de los diez experimentos y se aplica la prueba *ji*-cuadrado a los totales o supermuestra. La hipótesis nula, sigue siendo que la proporción de las semillas es de tres a uno. Sin embargo, se está haciendo la suposición adicional de que las diez muestras provienen de la misma población. Si este supuesto es correcto, se dice que las muestras son homogéneas; si el supuesto es falso, se dice que son heterogéneas y el análisis *ji*-cuadrado para las muestras combinadas no es justificable.

Es importante, entonces, realizar la prueba de heterogeneidad de las muestras. Si las muestras son homogéneas, el valor de X^2 total deberá estar muy cercano al valor de X^2 combinado. Al contrario, una diferencia apreciable entre estos dos valores es señal de heterogeneidad en los resultados experimentales. En el ejemplo presente, X^2 total $= 5.30$ con 10 *gl*. mientras que X^2 combinado $= 0.83$ con 1 *gl*. La diferencia $X^2 = 4.47$ tiene 9 *gl* y es menor que el valor tabulado $16.9 = \chi^2_{0.05,9}$.

Las hipótesis para probar heterogeneidad son:

H_0: las muestras son homogéneas.

H_1: las muestras son heterogéneas.

La diferencia $X^2 = 4.47$ es no significativa y se puede inferir homogeneidad. Se concluye, entonces, que las muestras provienen de la misma población o de poblaciones con una proporción idéntica en la forma de las semillas. Esta prueba también se conoce como prueba *ji*-cuadrado de interacción, es una prueba aproximada y puede reemplazarse por la prueba *G*, que también es aditiva.

Al concluir que las muestras son homogéneas, se justifica combinarlas para realizar la prueba con una muestra más grande y mejorar su potencia. En este momento puede aplicarse la corrección por continuidad si los datos así lo ameritan ($n < 200$).

Algunos autores, como Sokal y Rohlf (1995) y Zar (1999), discuten, con ejemplos hipotéticos, el caso de combinar muestras heterogéneas que conducen a resultados erróneos, o el caso cuando ninguna muestra individual conduce a rechazar H_0 pero la X^2 total resulta significativa. Otra situación conducente a la integración o combinación de varias

tablas 2 × 2 es el caso de estudios estratificados. Dos variables como presión sanguínea y ejercicio físico pueden presentar asociación creciente con la edad, la raza o la zona geográfica de residencia. La metodología para manejar estas tablas es la combinación ponderada de los datos en una tabla global, la cual se analiza, para asociación, con pruebas específicas como la de Cochran o la de Mantel y Haenszel (Everitt, 1992 y Rosner, 1999).

Investigaciones que analizan objetivos diferentes a la independencia de las variables pueden generar tablas $r \times c$ pero con una estructura diferente. Un caso especial son las tablas $r \times c$ que resultan de observaciones hechas antes-después a los mismos sujetos. En ellas se prueba simetría u homogeneidad marginal en vez de asociación y, por consiguiente, se requieren pruebas *ji*-cuadrado modificadas. Para tablas 2 × 2 la prueba apropiada es la de McNemar y para tablas mayores la prueba de Stuart-Maxwell. Estas pruebas y otras similares como la de concordancia entre evaluadores o entre tipos de escalas, se encuentran en Wickens (1989), Everitt (1992) y Bishop y otros (1995).

El análisis de tablas de contingencia se ha generalizado a tablas tridimensionales y de dimensiones superiores. Sin embargo, los objetivos varían originando problemas conceptuales que no existen en tablas $r \times c$ o bidimensionales, a la vez que incrementan la complejidad tanto del análisis como de la interpretación de los resultados. En los últimos treinta años se ha expandido la teoría sobre tablas de contingencia multidimensionales y han surgido nuevas metodologías estadísticas para su análisis como en los modelos lineales generalizados. En particular, se tienen los modelos log-lineales (el mejor método para analizar tablas de contingencia), la regresión logística, el análisis de correlación y análisis de correspondencia, véase por ejemplo, Wickens (1989), Cox y Snell (1992) y Agresti (2002).

12.8 Medidas de asociación en tablas $r \times c$

En las secciones anteriores se presentaron las pruebas de homogeneidad y de asociación en tablas $1 \times k$ y $r \times c$, respectivamente. El valor X^2 permite detectar asociación entre variables o diferencia entre proporciones, pero no mide el grado de asociación ni estima la relación entre las proporciones.

Agresti (2002) analiza varias medidas descriptivas para comparar tales proporciones, entre otras la diferencia de proporciones, el riesgo relativo y la razón de disparidad. También menciona algunas medidas de asociación como el coeficiente τ de Goodman y Kruskall, el coeficiente de incertidumbre U de Theil y los coeficientes de contingencia de Pearson y de Cramér. Estos últimos tienen una similitud con el coeficiente de determinación en variables continuas.

Ahora bien, un coeficiente de contingencia es un índice de asociación o dependencia entre dos variables categóricas y mide la intensidad de la asociación cuando la prueba *ji*-cuadrado ha permitido concluir que las variables están asociadas. El coeficiente de contingencia de Pearson se define como:

$$CC = \sqrt{X^2 / \left(X^2 + n\right)}.$$

El valor mínimo de este coeficiente es cero cuando no hay asociación, mientras que su valor máximo depende del tamaño de la tabla, así:

$$CC_{\text{máx.}} = \sqrt{(a-1)/a}\text{, con } a = \text{mín }(r, c).$$

En el ejemplo 12.4 se encontró dependencia entre el grupo sanguíneo y el SMSI. Se calcula cuánta es la dependencia mediante:

$$CC = \sqrt{18.858/(18.858 + 270)} = 0.2555.$$

Para comparar varios CC de tablas diferentes, es preferible obtener el CC relativo. Así:

$$CC_{\text{relativo}} = \frac{CC \times 100}{CC_{\text{máx}}}.$$

Para el ejemplo 12.4 el CC máx es $\sqrt{1/2} = 0.707$, luego el CC relativo será 100(0.2555 / 0.707) = 36.1%.

El coeficiente de Cramér tiene la ventaja de que varía entre 0 y 1 para cualquier tabla $r \times c$. Su expresión es:

$$V^2 = \sqrt{X^2/nd}, \text{ con } d = \text{mín}\,[(r-1), (c-1)].$$

Para el ejemplo 12.4 se tiene $V^2 = \sqrt{18.858/270 \times 1} = 0.264.$

Estas medidas de asociación basadas en X^2 son simétricas y se consideran significativas cuando la prueba *ji*-cuadrado también lo es. Sin embargo, su uso no es frecuente dada la dificultad en su interpretación. La única interpretación razonable es que cada medida se aleja de cero en la misma proporción en que los datos se alejan de la independencia (Bishop y otros, 1995).

Medidas de asociación, predicción, información, concordancia y máxima correlación en tablas de contingencia se estudian en Wickens (1989), Bishop y otros (1995), Agresti (2002) y Fleiss y otros (2003).

Ejercicios

12.1 En la siguiente tabla se dan los resultados observados en plantas de una población F_2 de pimientos. Los colores se esperan en las proporciones 1:3:8:4. Establecer la bondad del ajuste (Panse y Sukhatme, 1963).

Fenotipo	*Observados*
Púrpura oscuro	65
Púrpura medio	203
Púrpura claro	563
Verde	269
Total	1 100

12.2 En un estudio del estado de la caries dental en niños de seis comunidades con niveles variables de fluoruro en el suministro del agua, se seleccionó una muestra de 125 niños, de cada una de las comunidades, y se les practicó un examen dental (Daniel, 2004).

Comunidad	A	B	C	D	E	F
Niños sin caries	38	8	30	44	64	32

a. ¿Son homogéneas, en las proporciones de niños sin caries, las seis poblaciones?
b. Comparar proporciones por pares y establecer las diferencias entre las comunidades.

12.3 Gómez y González (1991) entrevistaron a 100 personas con el fin de establecer asociación entre el tipo de envoltura de la carne y el aspecto visual —atractivo, indiferente, desagradable— que presenta cuando se vende al público. Las frecuencias obtenidas fueron:

Aspecto	*Icopor*	*Shopak*	*Biopak*	*Cry-o-vac*
Atractivo	20	16	2	2
Indiferente	5	6	8	9
Desagradable	0	3	15	14

a. Probar asociación entre aspecto visual y envoltura. Analizar proporciones.
b. Descomponer la tabla en tres subtablas: icopor contra shopak, biopak contra cry-o-vac e icopor + shopak contra biopak + cry-o-vac. ¿Qué conclusiones pueden obtenerse?
c. Calcular el grado de asociación mediante los coeficientes de contingencia.

12.4 Los datos de la tabla adjunta corresponden a un estudio sobre un procedimiento quirúrgico diseñado para mejorar el funcionamiento de ciertas articulaciones afectadas por una enfermedad; tal procedimiento fue aplicado en cinco hospitales (Box, Hunter y Hunter, 2002).

Situación	*Hospitales*				
	A	*B*	*C*	*D*	*E*
Sin mejoría	13	5	8	21	43
Con mejoría	18	10	36	56	29
Recuperados	16	16	35	51	10

a. ¿Muestran estos datos igual efectividad del procedimiento quirúrgico en los cinco hospitales?
b. El hospital E es el hospital Central. ¿Existe diferencia entre este y los demás hospitales juntos?
c. ¿Existen diferencias entre los cuatro hospitales no centrales?
d. Explicar los resultados anteriores y relacionarlos con el resultado obtenido en *a*.

12.5 En un estudio para comparar las frecuencias del tipo de sangre y de varias enfermedades, los números observados fueron (Snedecor y Cochran, 1984):

Grupo sanguíneo	*Úlcera péptica*	*Cáncer gástrico*	*Controles*
O	983	383	2892
A	679	416	2625
B	134	84	570

a. Probar asociación entre los grupos sanguíneos y la enfermedad.
b. Analizar porcentajes para cada enfermedad.
c. Probar asociación en la subtabla de cáncer gástrico contra controles.
d. Probar asociación entre úlcera péptica y cáncer gástrico combinado con controles y los grupos A y B, sin tener en cuenta los tipos O.
e. Probar asociación entre el grupo O y los grupos A y B combinados, y las enfermedades úlcera péptica y cáncer gástrico + controles.
f. Explique la siguiente conclusión: la dependencia que se obtuvo inicialmente se debe fundamentalmente a una proporción mayor del tipo O entre los pacientes con úlcera péptica.

12.6 Con referencia al ejemplo 12.5, García y Gutiérrez observaron los siguientes resultados en los caninos.

	Positivos	*Negativos*
Adultos	19	94
Cachorros	18	74

Determinar la asociación entre la edad y la presencia de dermatofitos en los caninos.

12.7 Dos grupos de doce animales de experimentación, uno vacunado y el otro sin vacunar, se expusieron a la infección de una enfermedad. De acuerdo con los resultados de la siguiente tabla, ¿puede considerarse efectiva la vacuna contra tal enfermedad?

	Muertos	*Vivos*
Vacunados	2	10
No vacunados	8	4

Aplicar la prueba exacta de Fisher para la hipótesis unilateral y comparar los dos resultados.

12.8 En el ejemplo 12.6 se encontró asociación para X^2 sin corregir, pero debería concluirse independencia con X^2 corregido. ¿Cuál debe ser la decisión apropiada en este caso? Explicar esa decisión con argumentos estadísticos.

12.9 Un reporte médico decía que de 110 personas examinadas, 22 tenían parásitos de malaria en la sangre. De estas, cinco tenían el bazo crecido; mientras que en las que no tenían parásitos de malaria, solo tres presentaban dicho estado.

a. ¿Puede concluirse asociación entre estos dos factores?
b. ¿Cuáles son las probabilidades observadas en esta prueba?

12.10 En un estudio sobre el comportamiento de hormigas, se limpiaron —de hormigas— veintiocho árboles de acacia, de dos especies A y B. mediante insecticida. Luego se trajeron dieciséis colonias de hormigas recogidas en árboles de la especie A y se colocaron a igual distancia de los árboles, permitiéndose la invasión. Se registró el número de árboles que fueron reinfestados por las hormigas —no más de una colonia invade un árbol—. ¿Puede afirmarse que las hormigas invaden la especie en la que han vivido antes? (Sokal y Rohlf, 1995).

Especie	*Invadidos*	*No invadidos*
A	13	2
B	3	10

12.11 En cierto cruzamiento de dos variedades de guisantes, se espera que la mitad de las semillas sean rugosas y la otra mitad tersas. Se repitió ocho veces el mismo experimento y los resultados obtenidos se muestran a continuación (Ostle, 1988):

Experimento:	1	2	3	4	5	6	7	8
N.º rugosas:	60	108	80	118	165	106	105	90
N.º tersas:	40	92	100	90	135	76	125	110

a. Probar heterogeneidad de los experimentos.
b. ¿Pueden combinarse los resultados y analizarlos como una sola muestra? Explicar la respuesta.

12.12 En un experimento de genética sobre moscas domésticas, se hicieron siete repeticiones —cruces— en siete progenies, y se esperaba proporción de descendencia de nueve tipo salvaje, N^+, a siete de ojos pálidos, ge. Los resultados obtenidos se muestran en la siguiente tabla (Sokal y Rohlf, 1995).

Progenie		1	2	3	4	5	6	7
Fenotipo	N^+	83	77	110	92	51	48	70
	ge	47	43	96	58	31	61	42

a. Probar heterogeneidad de los experimentos.
b. Explicar la significancia de X^2 combinado y X^2 total.

Referencias bibliográficas

Agresti, A. 1984. *Analysis of Ordinal Categorical Data.* New York: John Wiley and Sons.
___. 2002. *Categorical Data Analysis.* 2nd ed., New York: John Wiley and Sons.
Ander Egg, E. 2001. *Métodos y técnicas de investigación social. Acerca del conocimiento y del pensamiento científico.* Buenos Aires: Lumen (Paidós).
Anderson, T. W. Darling, D. A. 1952. *Asymptotic theory of certain goodness of fit criteria based on stochastic processes.* Ann. Math. Stat. 23: 193-212.
Anderson, V. L. McLean, R. A. 1974. *Design of Experiments.* New York: Marcel Dekker.
Anderson-Sprecher, R. 1994. *Model comparisons and R^2.* The American Statistician 48: 113-117.
Arenas, J. y otros. 1985. *Extracción de pectinas a partir de desechos cítricos.* (Tesis de grado). Medellín: Universidad Pontificia Bolivariana.
Arnau Gras, J. 1981. *Diseños experimentales en psicología y educación.* vol. 1, México: Trillas.
Aubad, M. M. 1989. *Comportamiento eléctrico de dos especies ícticas frente a la presencia de sustancias contaminantes en el medio.* (Tesis de grado). Medellín: Universidad de Antioquia.
Baggaley, A. R. 1965. *Intermediate Correlational Methods.* New York: John Wiley and Sons.
Barnard, G. A. 1999. *On alleged gains in power from lower p-values.* Statistics in Medicine 8: 1469-1477.
Barnett, V. Lewis, T. 1994. *Outliers in Statistical Data.* 3th ed., London: John Wiley and Sons.
Bartlett, M. S. 1947. The Use of Transformations. *Biometrics* 3: 39-53.
___. 1949. *Fitting a Straight Line when both Variables are Subject to error.* Biometrics 5: 207-212.
Bazigos, G. P. 1976. *Estadísticas aplicadas de pesca.* Doc. Téc. FAO. Pesca (135).
Berhardson, C. S. 1975. Type ı Error Rates when Multiple Comparison Procedures follow a Significant F Test of ANOVA. *Biometrics* 31: 229-232.
Beyer, W. H. 1966. *Handbook of Tables for Probability and Statistics.* Cleveland: Chemical Rubber Co.
Birkes, D. y Dodge, Y. 1993. *Alternative Methods of Regression.* New York: John Wiley and Sons.
Bishop, O. N. 1966. *Statistics for Biology.* Boston: Houghton Mifflin.
Bishop, Y. V. V., Fienberg, S. S. Holland, P. W. 1995. *Discrete Multivariate Analysis.* 12th printing Cambridge Mass: MIT Press.
Bland, J.M. y Altman, D. G. 1996 Statistics notes: Transforming data. *British Medical Journal* 312: (7033):770.
Box, G. E. P. 1954. *Some theorems of cuadratic forms applied in the Study of Analysis of Variance Problems. I Effect on Inequality of Variance in the One-way Classification.* Ann. Math. Stat. 25: 290-302.
___. Cox, D. R. 1964. *An Analysis of Tranformations.* J. R. Statis. Soc. B26: 211-252.
Box, G. E. P., Hunter, W. G. and Hunter, J. S. 2002. *Estadística para investigadores. Introducción al diseño de experimentos, análisis de datos y construcción de modelos.* Barcelona: Editorial Reverté, S. A.
___. 2005. *Statistics for Experimenters: Design, Innovation and Discovery.* 2nd ed., New York: John Wiley and Sons.
Brownlee, K. A. 1957. *The Principles of Experimental Design.* Industrial Quality Control 13: 12-20.
Bunge, M. 1983. *La investigación científica, su estrategia y su filosofía.* Barcelona: Ariel.

Burr, I. W. 1974. *Applied Statistical Methods*. New York: Academic Press.

Carmer, S. G. y Swanson, M. R. 1973. An Evaluation of ten Pairwise Múltiple Comparison by Monte Carlo Methods. *JASA Asocc*. 68: 66-74.

Chatterjee, A. S. y Price, B. 1999. *Regression Analysis by Example*. 3rd edition, New York: John Wiley and Sons.

Chow, S. C., Shao, J. y Wang, H. 2003. *Sample Size Calculations in Clinical Research*. Boca Raton, FL: Taylor and Francis Group.

Cochran, W. G. 1947. Some Consequences when the Assumptions for the Analysis of Variance are not Satisfied. *Biometrics* 3: 22-38.

___. 1954. *Some Methods for strengthening the Common χ^2 Test*. Biometrics 10: 417-451.

___. 1995. *Técnicas de Muestreo*. México: CECSA.

___, y Cox, G. M. 1995. *Diseños Experimentales*. 3.ª ed, México: Trillas.

Cox, D. R. 1992. *Planning of experiments*. 2nd ed., New York: John Wiley and Sons.

___, y Snell E. J. 1992. *Analysis of Binary Data*. 2nd ed., London: Chapman and Hall.

Crowder, M. J. y Hand, D. J. 1999. *Analysis of Repeated Measures*. Boca Ratón, Fl.: Chapman and Hall / CRC.

D'Agostino, R. B. 1971. An Obnibus Test of Normality for Moderate and Large Size Samples. *Biometrika*, 58: 341-348.

___, y Stephens, M. A., 1986. *Goodness-of-fit Techniques*. New York: Marcel Dekker.

Daniel, W. W. 2004. *Bioestadística*. 8.ª ed., México: Limusa.

Das, M. N. y Giri, N. C. 1986. *Design and Analysis of Experiments*. New Delhi: Wiley Eastern Ltd.

David, H. A. 1998. *First (?) occurrence of common terms in Probability and Statistics*. Amer. Statist. 52: 1, February 1998.

Dayton, C. M. y Schafer, W. D. 1973. Extended Tables of t and chi-square for Bonferroni Tests with unequal Error Allocation. *JASA Assoc*. 68: 78-83.

Desu, M. M. y Raghavarao, D. 2004. *Nonparametric Statical Methods for complete and censored data*. Boca Raton, FL: Chapman and Hall / CRC.

Diamond, W. J. 2001, *Practical Experiment Design for Engineers and Scientists*. 3th ed, New York: John Wiley and Sons.

Díaz, A. 1981. *Muestreo Estadístico*. Medellín: ACESTA y Secretaría de Educación Departamental.

___, y Gutiérrez, A. 1995. *Estadística General*. Medellín: Centro de Educación a Distancia y Extensión, Universidad de Antioquia.

Diggle, P. J., Liang, K. Y. y Zeger, S. L. 1994. *Analysis of Longitudinal Data*. New York: Oxford University Press, Inc.

Di Trocchio, F. 1998. *Las mentiras de la ciencia. ¿Por qué y cómo engañan los científicos?* Madrid: Alianza Editorial.

Dixon, W. J. 1950. Analysis of Extreme Values. *Ann. Math. Statis*. 21: 488-506.

___, 1953. Processing Data for Outliers. *Biometrics* 22: 74-89.

Dobson, A. 2001, *An Introduction to Generalized Linear Models*. 2nd ed., Londres: Chapman and Hall.

Donner, A. y Eliasziw, M. 1992. A goodness of fit approach to inference procedure for the kappa statistic: C. I. construction, significance testing and sample size estimation. *Statistics in Medicine* 11: 511-519.

Draper, N. R. y Hunter, W. G. 1969. Transformations: some Examples revisited. *Technometrics* 11: 23-40.

Draper, N. R. y Smith, H. 1998. *Applied Regression Analysis*. 3th ed., New York: John Wiley and Sons.

Dudewicz, E. J. y Nishra, S. N. 1988. *Modern Mathematical Statistics*. Singapore: John Wiley and Sons.

Duncan, D. B. 1955. Multiple Range and Multiple F Tests. *Biometrics* 11: 1-42.

Dunn, O. J. 1964. Multiple Comparisons using Rank Sums. *Technometrics* 6: 241-252.

Dunnett, C. W. 1955. A Multiple Comparison Procedure for comparing several Treatments with a Control. *JASA* 50: 1096-1121.

___, 1970. Multiple Comparison Test. *Biometrics* 26: 139-141.

Dupont, W. D. 2002. *Statical Modeling for Medical Researchers. A Simple Introduction to the Analysis of Complex Data*. Cambridge, UK: Cambridge University Press.

Eisenhart, Ch. 1947. The Assumptions Underlying the Analysis of Variance. *Biometrics* 3: 1-21.

Escobar, L. D. y Urán, L. A. 1980. Introducción al método científico. *Actualidades Biológicas*, vol. 9: 57-61. Medellín: Universidad de Antioquia.

Everitt, B. S. 1992. *The Analysis of Contingency Tables*. 2nd ed., Londres: Chapman and Hall.

Fienberg, S. E. 1980. *The Analysis of cross-classified Categorical Data*. Cambridge, Mass.: MIT Press.

Finney, D. J. 1963. *Introduction to the Theory of Experimental Design.* Chicago: University of Chicago Press.
___, 1968. *Experimental Design and its Statistical Basis.* Chicago: University of Chicago Press.
___, 1978, *Statistical Method in Biological Assay.* London: Charles Griffin and Co.
Fisher, L. 1970. *Statistical Methods for Research Workers.* 14th ed., New York: Hafner Press.
Fisher, R. A. 1974. *The Design of Experiments.* 9th ed., London: Hafner Press.
___, y McDonald, J. 1978. *Fixed Effects Analysis of Variance.* New York: Academic Press.
Fleiss, J. L., Levin, B. and Cho Paik, M. 2003. *Statistical Methods for Rates and Proportions.* 3th ed., New York: John Wiley and Sons.
Folks, J. L. 1981. *Ideas of Statistics.* New York: John Wiley and Sons.
Friedman, M. 1937. *The Use of Ranks to avoid the Assumptions of Normality implicit in the Analysis of Variance.* JASA 32: 675-701.
Gardner, E. J. Simmons, M. J. y Snustad, D. P. 1991. *Principles of Genetics.* New York: John Wiley and Sons.
Gaviria, O. D., Hincapié, L. E. y Ramírez, J. 1989. *Influencia de la frecuencia de corte sobre la producción y composición química de cuatro variedades de avena sativa L.* (Tesis de grado). Medellín: Universidad de Antioquia.
George, J., Viña, G., Ramírez, A. y Mojica, J. I. 1991. *Manual de métodos de monitoreo biológico con aplicación en la industria del petróleo.* Bogotá: Ecopetrol.
Gibbons, J. D. 1993. *Nonparametric Statistics: An Introduction.* Newbury Park, Ca.: Sage Publications.
Gill, J. L. 1977. *Multiple Comparisons of Means when Variance is not Homogeneous.* J. Dairy Sci. 60: 444-449.
___, 1978. *Design and Analysis of Experiments in the Animal and Medical Sciences.* Ames, Iowa: The Iowa State University Press.
Giraldo, A. J. y Vargas, C. C. 1991. *Aspectos de la historia de vida de la sardina Jenskinsia lamprotaenia y su implicación social en el archipiélago de las Islas del Rosario* (Tesis de grado). Medellín: Universidad de Antioquia.
Gómez, H. 1989. *Estadística experimental con aplicaciones a las ciencias agrícolas.* Medellín: Universidad Nacional de Colombia.
Gómez, J. R. 1992. *Crecimiento fetal estandarizado por ecobiometría en una muestra obstrética de la ciudad de Medellín y su utilidad en el diagnóstico prenatal.* (Tesis de grado). Medellín: Universidad de Antioquia.
Gómez, L. M. y González, M. L. 1991. *Evaluación técnica de diferentes materiales de empaque como sustituto de la bandeja de poliestireno espumado utilizado para carne comercial de res y cerdo.* (Tesis de grado). Medellín: Corporación Universitaria Lasallista.
González, M. W. 1991. *Aspectos bioecológicos de los camarones de agua dulce* Macrobracium acantums *y* M. carcinus en la región de Urabá (Tesis de grado). Medellín: Universidad de Antioquia.
Goodall. C. 1990. *A survey of smoothing techniques. In: Modern Methods of Data Analysis* (Fox, J. y Long, J. S. eds.), pp. 126-176. Newbury Park, Ca.: Sage Publications.
Gosset, W. S. 1937. Comparison between Balanced and Random Arrangements of Field Plots. *Biometrika* 29: 368-379.
Govindarajulu, Z. 2001. *Statistical Techniques in Bioassay.* 2nd ed., Basilea: Karger Publishers.
Greenberg, B. G. 1951 Why Randomize? *Biometrics* 7: 309-322.
Grimes, D. y Schulz, K. 2002. Epidemiology Series. *The Lancet,* vol. 359, Nos. 9300-9310.
Grizzle, J. E. 1967. *Continuity correction in the* χ^2 *test for* 2×2 *tables.* Amer. Statist. 21: 28-32.
Grubbs, F. E. 1969. *Procedures for Detecting Outlying Observations in Samples.* Technometrics 11: 1-21.
Harrel, Jr. F. E. 2001. *Regression Modeling Strategies.* New York: Springer-Verlag.
Hastie, T. J. y Tibshirani, R. J. 1999. *Generalized Additive Models.* Boca ratón, Fl.: Chapman and Hall / CRC.
Hicks, C. R. y Turner, K. V. 1999. *Fundamental Concepts in the Design of Experiments.* 3th ed., New York: Oxford University Press.
Hinkelmann, K. and Kemptorne, O. 1994. *Design and Analysis of Experiments.* Vol .1: Introduction to Experimental Design. New York: John Wiley and Sons.
Hoaglin, D. C., Mosteller F. and Tukey, J. W. 2000. *Understanding Robust and Exploratory Data Analysis.* New York: John Wiley and Sons.
Hocking, R. R. 1996. *Methods and Applications of Linear Models.* New York: John Wiley and Sons.
Hollander, M. and Wolfe, D. A. 1999. *Nonparametric Statistical Methods.* 2nd ed., New York: John Wiley and Sons.

Hosmer, D. W. y Lemeshow, S. 2000. *Applied Logistic Regression.* 2nd ed., New York: John Wiley and Sons.

Hsu, J. C. 1996. *Multiple Comparisons.* Londres: Chapman and Hall.

Jobson, J. D. 1991, *Applied Multivariate data Analysis. I: Regression and Experimental Design.* New York: Springer-Verlag Inc.

John, J. A. y Quenouille, M. H. 1977. *Experiments: Design and Analysis.* London: Charles Griffin and Co.

Jones, D. 1984. Use, misuse, and role of multiple-comparison procedures in ecological and agricultural entomology. *Environmental Entomology* 13: 635-649.

Kimball, A. W. 1954. Short-cut Formulas for the Exact Partition of χ^2 in Contingency Tables. *Biometrics* 10: 452-458.

Kish, L. 1995. *Survey Sampling.* New York: John Wiley and Sons.

Kokoska, S y Nevison C. H. 1989. *Statistical tables and Formulae.* New York: Springer-Verlag Inc.

Koopmans, L. H. 1987. *An Introduction to Contemporary Statistics,* 2nd ed., Boston: Duxbury Press.

Kramer, C. Y. 1956. Extension of Multiple Range Test to Group Means with Unequal Numbers of Replications. *Biometrics* 12: 307-310.

Kruskal, W. H. y Wallis, W. A. 1952. *Use of Ranks in One-criterion Analysis of Variance.* JASA 47: 583-621 y 48: 907-911. (errata).

Lee, E. T. y Wang J. W. 2003. *Statistical Methods for Survival Data Analysis.* 3th ed., Belmont, Ca.: Lifetime Learning Publications.

Legendre, P. y Legendre, L. 1998. *Numerical Ecology.* 2nd ed., Amsterdam: Elsevier Science.

Lehmann, E. L. y D'Abrera, H. J. M. 1975. *Nonparametric Statistics Methods Based on Ranks.* San Francisco: Holden-Day.

Li, C. C. 1969a. *Introducción a la estadística experimental.* Barcelona: Omega S. A.

___, 1969b. *Genética humana. Principios y métodos.* Barcelona: Omega S. A.

Lindgren, B. W. 1976. *Statistical Theory.* New York: MacMillan Pu. Co.

Lindman, H. R. 1992. *Analysis of Variance in Experimental Design.* New York: Springer-Verlag.

Lohr, S. 1999. *Sampling: Design and Analysis.* Boston: Duxubury Press.

Ludwing, J. A. y Reynolds, J. F. 1988. *Statistical Ecology.* New York: John Wiley and Sons.

Madrigal, A. y Serna, M. A. 1978. Estudio comparativo de poblaciones de aves en plantaciones homogéneas y bosques secundarios. *Actualidades Biológicas,* vol. 6: 58-69. Medellín: Universidad de Antioquia.

Manotas, R., Montoya, F. y Sánchez, F. 1996. *Metodología de la Investigación en Ciencias de la Salud.* 3ra. ed., Medellín: Facultad de Medicina, Universidad de Antioquia.

Marks, R. 1982. *Designing a Research Project.* New York: Van Nostrand Reinhold Co.

Maxwell, A. W. 1967. *Analyzing Qualitative Data.* London: Methuen Inc.

McGarigal, K. Cushman, S. y Stafford, S. 2000. *Multivariate Statistics for Wildlife and Ecology Research.* New York: Springer-Verlag Inc.

Mehta, C. R. y Patel, N. R. 1986. *A hybrid algorithm for Fisher's exact test on unordered $r \times c$ contingency tables.* Communications in Statistics 15, 387-403.

Mendenhall, W. 1968. *The Design and Analysis o f Experiments.* Belmont, Ca: Wardsworth Pu. Co.

Miller, I. y Freund J. E. 1965. *Probability and Statistics for Engineers.* New Jersey: Prentice Hall.

Miller, R. G. Jr. 1981. *Simultaneous Statistical Inference.* 2nd ed, New York: Springer-Verlag.

___, 1986. *Beyond ANOVA, Basics of Applied Statistics.* New York: John Wiley and Sons.

Montgomery, D. C. 2007. *Diseño y Análisis de Experimentos.* 2.ª ed., México: Limusa Wiley.

Moreno, M. N.; García, D.; Sierra, L. I.; Calderón, M. A.; Patiño, P. y Díaz, A. 1992. Estudio de la Migración y de la actividad metabólica de los polimorfonucleares neutrófilos en pacientes con periodontitis juvenil localizada. *Iatreia* 5: 171-179. Medellín: Universidad de Antioquia.

Motulsky, H. 1995. *Intuitive Biostatistics.* New York: Oxford University Press. Inc.

Natrella, M. G. 2005. *Experimental Statistics.* New York: Dover Publications.

Neter, J., Kutner, M. H. Machtscheim, C. J. y Wassermann, W. 1996. *Applied Linear Statistical Models.* 4th ed., Homewood IL: Irwin.

Ortega, O. F. 1992. *Ciclo de vida, morfología, hábitos alimenticios, comportamiento y ecología de* Micrathyria Ocellata *(Calvert) (Odonata: libellulidae) depredador de larvas de mosquitos.* (Tesis de grado). Medellín: Universidad de Antioquia.

Ortiz, E. 1991. *Ensayos de sobrevivencia, levante y dietas alimenticias en larvas de* Colossoma Macropomum *(Cachama negra).* (Tesis de grado). Medellín: Universidad de Antioquia.

Ospina, D. 2001, *Introducción al muestreo*. Bogotá: Universidad Nacional de Colombia.
Ostle, B. 1988. *Estadística Aplicada*. 7.ª ed., México: Limusa.
Panse, V. G. y Sukhatme, P. V. 1963. *Métodos estadísticos para investigadores agrícolas*. 2.ª ed., México: Fondo de Cultura Económica.
Pareja, H. 1985. *Ecología y algunos aspectos del comportamiento en* Diplura *sp. (Aracneae: Dipluride)*. (Tesis de grado). Medellín: Universidad de Antioquia.
Paz, N. 1980. Estudio biológico en *Guzmania sp. (Bromeliaceae)* y *Heliconia Bihai (Musaceae)* en un bosque pluvial del Chocó. *Actualidades Biológicas* vol. 9: 31-50. Medellín: Universidad de Antioquia.
Pococh, S. J. 1987. *Clinical Trials: A Practical Approach*. New York: John Wiley and Sons.
Quinn, G. P. y Keough, M. J. 2003. *Experimental Design and Data Analysis for Biologists*. New York: Cambridge University Press.
Raghavarao, D. 1971. *Constructions and Combinational Problems in Design of Experiments*. New York: John Wiley and Sons.
Ramírez, A. 1988. *Lineamientos y estadísticas para estudios biológicos de impacto ambiental*, Bogotá: Publicaciones, Inderena.
Ramírez, J. J. y Machado, T. 1982. Influencia de la precipitación y los ortofosfatos sobre el fitoplancton de la represa La Fe. *Actualidades Biológicas* vol. 11: 3-21. Medellín: Universidad de Antioquia.
Rojas D., Ríos, O., Atehortúa, L., Díaz, A.; Echeverry, B., López, J., Botero B. y Valencia, C. 1990. *Captación de cinco metales pesados por* Pistia stratiotes *L. (Lechuga de agua)*. Medellín: Informe Científico presentado a Colciencias y a la Universidad de Antioquia.
Rosner, B. 1999. *Fundamentals of Biostatistics*. 5th ed., Boston: Duxbury Press.
Royston, J. P. 1982. An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics* 31: 115-124.
Saville, D. J. 1990. *Multiple comparison procedures. The practical solution*. American Statistician 44: 174-180.
Scheffé, H. 1999. *The Analysis of Variance*. New York: John Wiley and Sons.
Schefler, W. 1981. *Bioestadística*. Reading, Massachusetts. Fondo Educativo Interamericano.
Scheiner, S. M. y Gurevitch, J eds 2001. *Design and analysis of Ecological Experiments*. 2nd ed, New York: Oxford University Press., Inc.
Schlesselman, J. J. 1982. *Case-Control Studies. Design, Conduct, Analysis*. New York: Oxford University Press.
Serna, M. A. 1977. Distribución altitudinal y ecología de dos especies de aves simpátricas. *Actualidades biológicas*, vol. 6: 28-49. Medellín: Universidad de Antioquia.
Shapiro, S. S. y Francia, R. S. 1972. An approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, vol. 67: 215-216.
Shapiro, S. S. y Wilk, M. B. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52: 591-611.
Sheskin, D. J. 2000. *Handbook of Parametric and Nonparametric Statistical Procedures*. 2nd ed., Boca Ratón, Fl.: Chapman and Hall.
Simonoff, J. S. 1996. *Smoothing Methods in Statistics*. New York: Springer-Verlag.
Snedecor, G. W. y Cochran, W. G. 1984. *Métodos Estadísticos*. México: CECSA.
Snee, R. D. 1974. *Graphical Display of Two-way Contingency Tables*. Amer. Statist. 28: 9-12.
Sokal, R. R. y Rohlf, F. J. 1980. *Introducción a la Bioestadística*. Barcelona: Reverté.
___, 1995. *Biometry*. 3th ed., New York: W. H. Freeman and Co.
Spjøtvall, E. y Stoline, M. R. 1973. Extension of the T-Method of Multiple Comparison to include Cases with unequal Sample Sizes. *JASA* 68: 975-978.
Steel, R. G. D. y Torrie, J. H. 1997. *Bioestadística: principios y procedimientos*. México: McGraw-Hill.
Storm, L. E. 1962. *Nested Analysis of Variance: Review of Methods*. Metrika 5: 158-183.
Sukhatme, P. V. y Sukhatme, B. V. 1970. *Sampling theory of Survey with Applications*. Bombay: Asia Publishing House.
Thode, Jr., H. C. 2002. *Testing for Normality*. New York: Marcel Dekker.
Thomas, L. y Juanes, F. 1996. The importance of statistical power analysis: An Example from animal behaviour. *Animal Behaviour* 52: 856-859.
Thompson S. K. 2002. *Sampling*. 2nd ed., New York: John Wiley and Sons.
Tietjen, G. L y Moore, R. H. 1972. Some Grubbs-Type Statistics of the Detection of Several Outliers. *Technometrics*, 14: 583-597.

Trujillo F. J. y Pulido, H. 1986. *Metodología para el estudio de aguas de lavado ("leaching") de materiales sólidos y su impacto ambiental.* (Informe de investigación). Medellín: Universidad de Antioquia.

Tukey, J. W. 1949. One Degree of Freedom for non-additivity. *Biometrics* 5: 232-242.

___, 1977. *Exploratory Data Analysis.* Reading, Mass.: Addisson Wesley.

___, y McLaughlin, D. H. 1963. Less vulnerable confidence and significance procedures for location based on a single sample: Trimming / Winsorization. *Sankya.* Series A 25: 334-352.

Underwood, A. J. 2001. *Experiments in Ecology: Their logical design and interpretation using analysis of variance.* New York: Cambridge University Press.

Vallejo A. 1989. *Cultivo del rotífero Brachionus plicatilis O. F. Müller (*1786*) a nivel masivo en el laboratorio.* (Tesis de grado). Medellín: Universidad de Antioquia.

Velleman, P. F. y Hoaglin, D. C. 1981. *Applications, Basics and Computing of Exploratory Data Analysis.* Boston: Duxbury Press.

Wardlaw, A. C. 2000. *Practical Statistics for Experimental Biologists.* 2nd ed., New York: Wiley Chichester.

Weerahandi, S. 1996. *Exact Statistical Methods for Data Analysis.* New York: Springer-Verlag, Inc.

Wickens, T. D. 1989. *Multiway Contingency Tables Analysis for the Social Sciences.* New Jersey: Lawerence Evlbaum associates, Inc.

Wilcox, R. R. 1997. *Introduction to the Robust Estimation and Hypothesis Testing.* San Diego: Academic Press.

Youden, W. J. 1972. *Randomization and Experimentation.* Technometrics 14: 13-22.

Zar, J. H. 1999. *Biostatistical Analysis.* 4th ed., New Jersey: Prentice Hall, Inc.

Apéndices

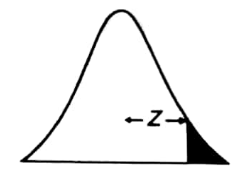
A . Tablas

Tabla A.1 Números aleatorios

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
1	48461	14952	72619	73689	52059	37086	60050	86192	67049	64739	1
2	76534	38149	49692	31366	52093	15422	20498	33901	10319	43397	2
3	70437	25861	38504	14752	23757	59660	67844	78815	23758	86814	3
4	59584	03370	42806	11393	71722	93804	09095	07856	55589	46020	4
5	04285	58554	16085	51555	27501	73883	33427	33343	45507	50063	5
6	77340	10412	69189	85171	29082	44785	83638	02583	96483	76553	6
7	59183	62687	91778	80354	23512	97219	65921	02035	59847	91403	7
8	91800	04281	39979	03927	82564	28777	59049	97532	54540	79472	8
9	12066	24817	81099	48940	69554	55925	48379	12866	51232	21580	9
10	69907	91751	53512	23748	65906	91385	84983	27915	48491	91068	10
11	80467	04873	54053	25955	48518	13815	37707	68687	15570	08890	11
12	78057	67835	28302	45048	56761	97725	58438	91528	24645	18544	12
13	05648	39387	78191	88415	60269	94880	58812	42931	71898	61534	13
14	22304	39246	01350	99451	61862	78688	30339	60222	74052	25740	14
15	61346	50269	67005	40442	33100	16742	61640	21046	31909	72641	15
16	66793	37696	27965	30459	91011	51426	31006	77468	61029	57108	16
17	86411	48809	36698	42453	83061	43769	39948	87031	30767	13953	17
18	62098	12825	81744	28882	27369	88183	65846	92545	09065	22655	18
19	68775	06261	54265	16203	23340	84750	16317	88686	86862	00879	19
20	52679	19595	13687	74872	89181	01939	18447	10787	76246	80072	20
21	84096	87152	20719	25215	04349	54434	72344	93008	83282	31670	21
22	63964	55937	21417	49944	38356	98404	14850	17994	17161	98981	22
23	31191	75131	72386	11689	95727	05414	88727	45583	22568	77700	23
24	30545	68523	29850	67833	05622	89975	79042	27142	99257	32349	24
25	52573	91001	52315	26430	54175	30122	31796	98842	37600	26025	25
26	16586	81842	01076	99414	31574	94719	34656	80018	86988	79234	26
27	81841	88481	61191	25013	30272	23388	22463	65774	10029	58376	27
28	43563	66829	72838	08074	57080	15446	11034	98143	74989	26885	28
29	19945	84193	57581	77252	85604	45412	43556	27518	90572	00563	29
30	79374	23796	16919	99691	80276	32818	62953	78831	54395	30705	30
31	48503	26615	43980	09810	38289	66679	73799	48418	12647	40044	31
32	32049	65541	37937	41105	70106	89706	40829	40789	59547	00783	32
33	18547	71562	95493	34112	76895	46766	96395	31718	48302	45893	33
34	03180	96742	61486	43305	34183	99605	67803	13491	09243	29557	34
35	94822	24738	67749	83748	59799	25210	31093	62925	72061	69991	35
36	34330	60599	85828	19152	68499	27977	35611	96240	62747	89529	36
37	43770	81537	59527	95674	76692	86420	69930	10020	72881	12532	37
38	56908	77192	50623	41215	14311	42834	80651	93750	59957	31211	38
39	32787	07189	80539	75927	75475	73965	11796	72140	48944	74156	39
40	52441	78392	11733	57703	29133	71164	55355	31006	25526	55790	40
41	22377	54723	18227	28449	04570	18882	00023	67101	06895	08915	41
42	18376	73460	88841	39602	34049	20589	05701	08249	74213	25220	42
43	53201	28610	87957	21497	64729	64983	71551	99016	87903	63875	43
44	34919	78901	59710	27396	02593	05665	11964	44134	00273	76358	44
45	33617	92159	21971	16901	57383	34262	41744	60891	57624	06962	45
46	70010	40964	98780	72418	52571	18415	64362	90636	38034	04909	46
47	19282	68447	35665	31530	59832	49181	21914	65742	89815	39231	47
48	91429	73328	13266	54898	68795	40948	80808	63887	89939	47938	48
49	97637	78393	33021	05867	86520	45363	43066	00988	64040	09803	49
50	95150	07625	05255	83254	93943	52325	93230	62668	79529	65964	50

Fuente: Sokal y Rohlf (1995).

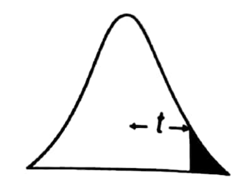
Tabla A.2 Distribución normal



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2336	0.2206	0.2177	0.2148
0.8	0.2119	0:2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	1.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Fuente: Box, Hunter y Hunter (2005).

Tabla A.3 Distribución *t* de Student



υ	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Fuente: Box, Hunter y Hunter (2005).

Tabla A.4 Coeficientes (a_{n-i+1}) para la prueba W de normalidad, y $n = 2(1)50$

i \ n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2		.0000	.1677	.2413	.2806	.3031	.3164	.3244	.3291
3				.0000	.0875	.1401	.1743	.1976	.2141
4						.0000	.0561	.0947	.1224
5								.0000	.0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	.3315	.3325	.3325	.3318	.3306	.3290	.3273	.3253	.3232	.3211
3	.2260	.2347	.2412	.2460	.2495	.2521	.2540	.2553	.2561	.2565
4	.1429	.1586	.1707	.1802	.1878	.1939	.1988	.2027	.2059	.2085
5	.0695	.0922	.1099	.1240	.1353	.1447	.1524	.1587	.1641	.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7			.0000	.0240	.0433	.0593	.0725	.0837	.0932	.1013
8					.0000	.0196	.0359	.0496	.0612	.0711
9							.0000	.0163	.0303	.0422
10									.0000	.0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	.3185	.3156	.3126	.3098	.3069	.3043	.3018	.2992	.2968	.2944
3	.2578	.2571	.2563	.2554	.2543	.2533	.2522	.2510	.2499	.2487
4	.2119	.2131	.2139	.2145	.2148	.2151	.2152	.2151	.2150	.2148
5	.1736	.1764	.1787	.1807	.1822	.1836	.1848	.1857	.1864	.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	.1092	.1150	.1201	.1245	.1283	.1316	.1346	.1372	.1395	.1415
8	.0804	.0878	.0941	.0997	.1046	.1089	.1128	.1162	.1192	.1219
9	.0530	.0618	.0696	.0764	.0823	.0876	.0923	.0965	.1002	.1036
10	.0263	.0368	.0459	.0539	.0610	.0672	.0728	.0778	.0822	.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12			.0000	.0107	.0200	.0284	.0358	.0424	.0483	.0537
13					.0000	.0094	.0178	.0253	.0320	.0381
14							.0000	.0084	.0159	.0227
15									.0000	.0076

Tabla A.4 (continuación)

i \ n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	.2921	.2898	.2876	.2854	.2834	.2813	.2794	.2774	.2755	.2737
3	.2475	.2463	.2451	.2439	.2427	.2415	.2403	.2391	.2380	.2368
4	.2145	.2141	.2137	.2132	.2127	.2121	.2116	.2110	.2104	.2098
5	.1874	.1878	.1880	.1882	.1883	.1883	.1883	.1881	.1880	.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	.1433	.1449	.1463	.1475	.1487	.1496	.1505	.1513	.1520	.1526
8	.1243	.1265	.1284	.1301	.1317	.1331	.1344	.1356	.1366	.1376
9	.1066	.1093	.1118	.1140	.1160	.1179	.1196	.1211	.1225	.1237
10	.0899	.0931	.0961	.0988	.1013	.1036	.1056	.1075	.1092	.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	.0585	.0629	.0669	.0706	.0739	.0770	.0798	.0824	.0848	.0870
13	.0435	.0485	.0530	.0572	.0610 .	.0645	.0677	.0706	.0733	.0759
14	.0289	.0344	.0395	.0441	.0484	.0523	.0559	.0592	.0622	.0651
15	.0144	.0206	.0262	.0314	.0361	.0404	.0444	.0481	.0515	.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17			.0000	.0062	.0119	.0172	.0220	.0264	.0305	.0343
18					.0000	.0057	.0110	.0158	.0203	.0244
19							.0000	.0053	.0101	.0146
20									.0000	.0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	.2719	.2701	.2684	.2667	.2651	.2635	.2620	.2604	.2589	.2574
3	.2357	.2345	.2334	.2323	.2313	.2302	.2291	.2281	.2271	.2260
4	.2091	.2085	.2078	.2072	.2065	.2058	.2052	.2045	.2038	.2032
5	.1876	.1874	.1871	.1868	.1865	.1862	.1859	.1855	.1851	.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	.1531	.1535	.1539	.1542	.1545	.1548	.1550	.1551	.1553	.1554
8	.1384	.1392	.1398	.1405	.1410	.1415	.1420	.1423	.1427	.1430
9	.1249	.1259	.1269	.1278	.1286	.1293	.1300	.1306	.1312	.1317
10	.1123	.1136	.1149	.1160	.1170	.1180	.1189	.1197	.1205	.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	.0891	.0909	.0927	.0943	.0959	.0972	.0986	.0998	.1010	.1020
13	.0782	.0804	.0824	.0842	.0860	.0876	.0892	.0906	.0919	.0932
14	.0677	.0701	.0724	.0745	.0765	.0783	.0801	.0817	.0832	.0846
15	.0575	.0602	.0628	.0651	.0673	.0694	.0713	.0731	.0748	.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	.0379	.0411	.0442	.0471	.0497	.0522	.0546	.0568	.0588	.0608
18	.0283	.0318	.0352	.0383	.0412	.0439	.0465	.0489	.0511	.0532
19	.0188	.0227	.0263	.0296	.0328	.0357	.0385	.0411	.0436	.0459
20	.0094	.0136	.0175	.0211	.0245	.0277	.0307	.0335	.0361	.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22			.0000	.0042	.0081	.0118	.0153	.0185	.0215	.0244
23					.0000	.0039	.0076	.0111	.0143	.0174
24							.0000	.0037	.0071	.0104
25									.0000	.0035

Fuente: Anderson y McLean (1974).

Tabla A.5 Valores porcentuales de la prueba W para $n = 3(1)50$

n \ α	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	.687	.707	.748	.792	.935	.987	.992	.996	.997
5	.686	.715	.762	.806	.927	.979	.986	.991	.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	.730	.760	.803	.838	.928	.972	.979	.985	.988
8	.749	.778	.818	.851	.932	.972	.978	.984	.987
9	.764	.791	.829	.859	.935	.972	.978	.984	.986
10	.781	.806	.842	.869	.938	.972	.978	.983	.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	.805	.828	.859	.883	.943	.973	.979	.984	.986
13	.814	.837	.866	.889	.945	.974	.979	.984	.986
14	.825	.846	.874	.895	.947	.975	.980	.984	.986
15	.835 .	.855	.881	.901	.950	.975	.980	.984	.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	.851	.869	.892	.910	.954	.977	.981	.985	.987
18	.858	.874	.897	.914	.956	.978	.982	.986	.988
19	.863	.879	.901	.917	.957	.978	.982	.986	.988
20	.868	.884	.905	.920	.959	.979	.983	.986	.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	.878	.892	.911	.926	.961	.980	.984	.987	.989
23	.881	.895	.914	.928	.962	.981	.984	.987	.989
24	.884	.898	.916	.930	.963	.981	.984	.987	.989
25	.888	.901	.918	.931	.964	.981	.985	.988	.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	.894	.906	.923	.935	.965	.982	.985	.988	.990
28	.896	.908	.924	.936	.966	.982	.985	.988	.990
29	.898	.910	.926	.937	.966	.982	.985	.988	.990
30	.900	.912	.927	.939	.967	.983	.985	.988	.900
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	.904	.915	.930	.941	.968	.983	.986	.988	.990
33	.906	.917	.931	.942	.968	.983	.986	.989	.990
34	.908	.919	.933	.943	.969	.983	.986	.989	.990
35	.910	.920	.934	.944	.969	.984	.986	.989	.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	.914	.924	.936	.946	.970	.984	.987	.989	.990
38	.916	.925	.938	.947	.971	.984	.987	.989	.990
39	.917	.927	.939	.948	.971	.984	.987	.989	.991
40	.919	.928	.940	.949	.972	.985	.987	.989	.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	.922	.930	.942	.951	.972	.985	.987	.989	.991
43	.923	.932	.943	.951	.973	.985	.987	.990	.991
44	.924	.933	.944	.952	.973	.985	.987	.990	.991
45	.926	.934	.945	.953	.973	.985	.988	.990	.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	.928	.936	.946	.954	.974	.985	.988	.990	.991
48	.929	.937	.947	.954	.974	.985	.988	.990	.991
49	.929	.937	.947	.955	.974	.985	.988	.990	.991
50	.930	.938	.947	.955	.974	.985	.988	.990	.991

Fuente: Andereon y McLean (1974).

Tabla A.6 Distribución F para $\alpha = 0.10$

υ_2 \ υ_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
									$\alpha = 0.10$										
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.23	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.95	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1,67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.96	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Tabla A.6 (continuación)

$\alpha = 0.05$

υ_2 \ υ_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.4	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	18.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.23	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Tabla A.6 (continuación)

$\alpha = 0.01$

$\upsilon_2 \backslash \upsilon_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	2.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Tabla A.6 (continuación)

$F(\alpha = 0.005)$

υ_2 \ υ_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198.5	199.0	199.2	199.2	199,3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199,5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.531	2.40	12.27	12.14
6	18.63	15.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.79	4.65	4.55	4.44	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.25
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.29
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.24
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	2.09	1.95	1.87	1.75	1.61	1.43
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

Fuente: W. H. Beyer, 1966. *Handbook of Tables for Probability and Statistics*. The Chemical Rubber Co. Cleveland.

Tabla A.7 Distribución *ji*-cuadrado, χ^2

υ	*0.995*	*0.990*	*0.975*	*0.950*	*0.900*	*0.750*	*0.500*	*0,250*	*0.100*	*0.050*	*0.025*	*0.010*	*0.005*	*0.001*
1					0.02	0.10	0.45	1.32	2.71	3.84	"5.02	6.63	7.88	10.83
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.73	40.79
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.80	55.76	59.34	63.69	66.77	73.40
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22	112.32
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	113.14	118.14	124.12	128.30	137.21
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17	149.45

Fuente: Anderson y McLean (1974).

Tabla A.8 Distribución de $F_{máx}$

υ	α	*K* = 3	4	5	6	7	8	9	10	12
2	.25	15.4	25.1	35.5	46.8	58.5	70.8	83.6	96.7	124
	.10	42.5	69.1	98.2	129	162	196	231	267	342
	.05	87.5	142	202	266	333	403	475	550	704
	.01	448	729	1036	1362	1705	2063	2432	2813	3605
3	.25	8.13	11.7	15.1	18.5	21.9	25.2	28.4	31.6	37.8
	.10	16.8	24.0	30.9	37,7	44.4	50.9	57.4	63.7	76.1
	.05	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	124
	.01	85	120	151	184	216	249	281	310	361
4	.25	5.79	7.81	9.67	11.4	13.1	14.7	16.2	17.6	20.4
	.10	10.4	13.9	17.1	20.1	22.9	25.6	28.1	30.6	35.3
	.05	15.5	20.6	25.2	29.5	33.6	37.5	41.1	44.6	51.4
	.01	37	49	59	69	79	89	97	106	120
5	.25	4.66	6.05	7.28	8.41	9.46	10.4	11.4	12.2	13.9
	.10	7,68	9.86	11.8	13.5	15.2	16.7	18.1	19.4	21.9
	.05	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	29.9
	.01	22	28	33	38	42	46	50	54	60
6	.25	4.00	5.05	5.97	6.78	7.53	8.22	8.87	9.48	10.6
	.10	6.23	7.78	9.11	10.3	11.4	12.4	13.3	14.2	15.8
	.05	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	20.7
	.01	15.5	19.1	22	25	27	30	32	34	37
7	.25	3.57	4.11	5.14	5.77	6.35	6.88	7.37	7.82	8.66
	.10	5.32	6.52	7.52	8.41	9.20	9.93	10.6	11.2	12.4
	.05	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.8
	.01	12.1	14.5	16.5	18.4	20	22	23	24	27
8	.25	3.26	3.97	4.57	5.09	5.55	5.98	6.37	6.73	7.40
	.10	4.71	5.68	6.48	7.18	7.80	8.36	8.88	9.36	10.2
	.05	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.7
	.01	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	21
9	.25	3.03	3.64	4.15	4.59	4.98	5.34	5.66	5.96	6.51
	.10	4.26	5.07	5.74	6.32	6.82	7.28	7.70	8.09	8.78
	.05	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.7
	.01	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.6
10	.25	2.85	3.39	3.83	4.21	4.55	4.86	5.13	5.39	5.85
	.10	3.93	4.63	5.19	5.68	6.11	6.49	6.84	7.16	7.74
	.05	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.34
	.01	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.9
12	.25	2.58	3.02	3.38	3.68	3.95	4.18	4.40	4.60	4.95
	.10	3.45	4.00	4.44	4.81	5.13	5.42	5.68	5.92	6.35
	.05	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.48
	.01	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.6
15	.25	2.32	2.67	2.95	3.18	3.38	3.56	3.72	3.87	4.13
	.10	3.00	3.41	3.74	4.02	4.25	4.46	4.65	4.82	5.13
	.05	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.93
	.01	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7,5	8.0
20	.25	2.07	2.33	2.53	2.70	2.85	2.98	3.09	3.20	3.38
	.10	2.57	2.87	3.10	3.29	3.46	3.60	3.73	3.85	4.06
	.05	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.59
	.01	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.9
30	.25	1.80	1.98	2.12	2.24	2.34	2.42	2.49	2.56	2.68
	.10	2.14	2.34	2.50	2.62	2.73	2.82	2.90	2.97	3.10
	.05	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.39
	.01	3.00	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.2
60	.25	1.51	1.62	1.70	1.76	1.81	1.86	1.90	1.93	2.00
	.10	1.71	1.82	1.90	1.96	2.02	2.07	2.11	2.14	2.21
	.05	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.36
	.01	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7
∞		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Fuente: Gill (1978).

Tabla A.9 Valores porcentuales para la prueba *Q* de Burr y Foster, para igual número de grados de libertad υ y para *k* muestras

k	υ = 1		υ = 2		υ = 3		υ = 4	
	.99	.999	.99	.999	.99	.999	.99	.999
3	*	*	.863	*	.757	.919	.684	.828
4	.920	*	.720	.898	.605	.754	.549	.675
5	.828	*	.608	.773	.512	.644	.443	.552
6	.744	.949	.539	.690	.430	.546	.369	.461
7	.671	.865	.469	.606	.372	.471	.318	.394
8	.609	.793	.412	.537	.325	.411	.276	.342
9	.576	.750	.371	.481	.287	.363	.244	.300
10	.528	.694	.333	.433	.257	324	.218	.267
12	.448	.598	.276	.358	.211	.265	.179	.217
14	.39T	.522	.234	.303	.178	.222	.151	.181
15	.365	.490	.217	.280	.165	.205	.140	.167
16	.343	.460	.202	.261	.154	.190	.130.	.155
18	.304	.409	.178	.228	.135	.165	.114	.135
20	.273	.367	.158	.202	.120	.146	.101	.119
22	.246	.332	.142	.180	.108	.130	.090	.106
24	.224	.302	.129	.162	.098	.117	.082	.096
26	.206	.276	.118	.148	.090	.107	.075	.087
28	.190	.254	.108	.135	.082	.098	.069	.080
30	.176	.234	.100	.124	.075	.090	.064	.074
32	.163	.218	.093	.115	.070	.083	.060	.068
36	.143	.189	.082	.100	.062	.072	.052	.060
40	.127	.167	.072	.088	.055	.064	.047	.053
45	.111	.145	.063	.076	.048	.055	.041	.046
50	.098	.127	.056	.067	.043	.049	.037	.041
60	.080	.102	.045	.053	.035	.039	.030	.033
64	.074	.094	.042	.049	.033	.037	.028	.031

* Dado que $Q \geq 1$, estas entradas se omiten.

Tabla A.9 (continuación)

k	υ = 5		υ = 6		υ = 8		υ = 10	
	.99	*.999*	*.99*	*.999*	*.99*	*.999*	*.99*	*.999*
3	.631	.760	.593	.708	.539	.633	.512	.596
4	.498	.608	.461	.558	.413	.490	.383	.446
5	.399	.490	.368	.446	.328	.388	.303	.351
6	.334	.407	.307	.368	.271	.318	.250	.288
7	.284	.345	.261	.311	.230	.268	.212	.242
8	.246	.298	.226	.268	.199	.231	.184	.209
9	.217	.261	.199	.235	.176	.202	.162	.183
10	.194	.232	.178	.208	.157	.179	.145	.163
15	.123	.145	.113	.131	.101	.113	.094	.103
20	.090	.104	.083	.094	.074	.082	.069	.075
30	.058	.065	.053	.059	.048	.052	.045	.048
40	.042	.047	.039	.043	.035	.038	.033	.035
50	.033	.036	.031	.033	.028	.030	.026	.028
60	.027	.029	.025	.027	.023	.024	.022	.023

Tabla A.9 (continuación)

k	υ = 12		υ = 14		υ = 16		υ = 20	
	.99	*.999*	*.99*	*.999*	*.99*	*.999*	*.99*	*.999*
3	.486	.558	.466	.530	.451	.508	.429	.476
4	.362	.415	.347	.393	.335	.375	.319	.351
5	.287	.326	.275	.308	.265	.295	.252	.276
6	.236	.267	.227	.253	.219	.242	.209	.226
7	.201	.225	.192	.213	.186	.204	.178	.191
8	.174	.194	.167	.184	.162	.176	.154	.166
9	.154	.170	.148	.162	.143	.155	.136	.146
10	.137	.152	.132	.144	.128	.138	.122	.130
15	.089	.097	.086	.092	.083	.089	.080	.084
20	.066	.070	.063	.067	.062	.065	.059	.062
30	.043	.045	.042	.043	.040	.042	.039	.040
40	.032	.033	.031	.032	.030	.031	.029	.030
50	.025	.026	.024	.025	.024	.025	.023	.024
60	.021	.022	.020	.021	.020	.020	.019	.020

Fuente: Anderson y McLean (1974).

Tabla A.10 Valores de t_B para el intervalo de Bonferroni

m	α	υ = 10	12	14	16	18	20	22	24	26	v = 30	35	40	50	60	80	100	250	500
1	.05	1.812	1.782	1.761	1.746	1.734	1.725	1.717	1.711	1.706	1.697	1.690	1.684	1.676	1.671	1.664	1.660	1.651	1.648
	.025	2.228	2.179	2.145	2.120	2.101	2.086	2.074	2.064	2.056	2.042	2.030	2.021	2.009	2.000	1.990	1.984	1.970	1.965
	.01	2.764	2.681	2.624	2.583	2.552	2.528	2.508	2.492	2.479	2.457	2.438	2.423	2.403	2.390	2.374	2.364	2.341	2.334
	.005	3.169	3.055	2.977	2.921	2.878	2.845	2.819	2.797	2.779	2.750	2.724	2.704	2.678	2.660	2.639	2.626	2.596	2.586
2	.05	2.228	2.179	2.145	2.120	2.101	2.086	2.074	2.064	2.056	2.042	2.030	2.021	2.009	2.000	1.990	1.984	1.970	1.965
	.025	2.634	2.560	2.510	2.473	2.445	2.423	2.405	2.391	2.379	2.360	2.342	2.329	2.311	2.299	2.284	2.276	2.255	2.248
	.01	3.169	3.055	2.977	2.921	2.878	2.845	2.819	2.797	2.779	2.750	2.724	2.704	2.678	2.660	2.639	2.626	2.596	2.586
	.005	3.581	3.428	3.326	3.252	3.197	3.158	3.119	3.091	3.067	3.030	2.996	2.971	2.937	2.915	2.887	2.871	2.832	2.820
3	.05	2.468	2.403	2.360	2.324	2.304	2.285	2.270	2.258	2.247	2.231	2.215	2.204	2.188	2.178	2.165	2.158	2.140	2.134
	.025	2.870	2.779	2.718	2.673	2.639	2.613	2.591	2.574	2.559	2.536	2.515	2.499	2.477	2.463	2.445	2.435	2.410	2.402
	.01	3.409	3.273	3.181	3.115	3.065	3.026	2.995	2.970	2.949	2.915	2.885	2.862	2.831	2.811	2.786	2.771	2.736	2.725
	.005	3.827	3.649	3.530	3.444	3.380	3.331	3.291	3.258	3.231	3.189	3.150	3.122	3.083	3.057	3.026	3.007	2.964	2.949
4	.05	2.634	2.560	2.510	2.473	2.445	2.428	2.405	2.391	2.379	2.360	2.342	2.329	2.311	2.299	2.284	2.276	2.255	2.248
	.025	3.038	2.934	2.864	2.813	2.775	2.744	2.720	2.700	2.684	2.657	2.633	2.616	2.591	2.575	2.555	2.544	2.516	2.507
	.01	3.581	3.428	3.326	3.252	3.197	3.153	3.119	3.091	3.067	3.030	2.996	2.971	2.937	2.915	2.887	2.871	2.832	2.820
	.005	4.005	3.807	3.675	3.581	3.510	3.455	3.412	3.376	3.346	3.300	3.258	3.227	3.184	3.156	3.122	3.102	3.054	3.039
5	.05	2.764	2.681	2.624	2.583	2.552	2.528	2.508	2.492	2.479	2.457	2.438	2.423	2.403	2.390	2.374	2.364	2.341	2.334
	.025	3.169	3.055	2.977	2.921	2.878	2.845	2.819	2.797	2.779	2.750	2.724	2.704	2.678	2.660	2.639	2.626	2.596	2.586
	.01	3.716	3.550	3.438	3.358	3.298	3.251	3.214	3.183	3.158	3.118	3.081	3.055	3.018	2.994	2.964	2.945	2.905	2.892
	.005	4.144	3.930	3.787	3.686	3.610	3.552	3.505	3.467	3.435	3.385	3.340	3.307	3.261	3.232	3.195	3.174	3.123	3.107
6	.05	2.870	2.779	2.718	2.673	2.639	2.613	2.591	2.574	2.559	2.536	2.515	2.499	2.477	2.463	2.445	2.435	2.410	2.402
	.025	3.277	3.153	3.069	3.008	2.963	2.927	2.899	2.875	2.856	2.825	2.797	2.776	2.747	2.729	2.705	2.692	2.659	2.649
	.01	3.827	3.649	3.530	3.444	3.380	3.331	3.291	3.258	3.231	3.189	3.150	3.122	3.083	3.057	3.026	3.007	2.964	2.949
	.005	4.259	4.031	3.880	3.773	3.692	3.630	3.581	3.540	3.507	3.454	3.407	3.372	3.324	3.293	3.254	3.232	3.179	3.161
7	.05	2.960	2.863	2.796	2.748	2.712	2.683	2.661	2.642	2.626	2.601	2.579	2.562	2.539	2.524	2.505	2.494	2.467	2.459
	.025	3.368	3.236	3.146	3.082	3.034	2.996	2.965	2.941	2.920	2.887	2.857	2.836	2.805	2.785	2.761	2.747	2.712	2.701
	.01	3.922	3.733	3.607	3.517	3.450	3.398	3.356	3.322	3.293	3.249	3.208	3.178	3.137	3.111	3.078	3.058	3.013	2.998
	.005	4.357	4.117	3.958	3.846	3.762	3.697	3.645	3.603	3.567	3.513	3.463	3.426	3.376	3.344	3.304	3.280	3.225	3.207
8	.05	3.038	2.934	2.864	2.813	2.775	2.744	2.720	2.700	2.684	2.657	2.633	2.616	2.591	2.575	2.555	2.544	2.516	2.507
	.025	3.448	3.308	3.214	3.146	3.095	3.055	3.023	2.997	2.975	2.941	2.910	2.887	2.855	2.834	2.809	2.793	2.758	2.745
	.01	4.005	3.807	3.675	3.581	3.510	3.455	3.412	3.376	3.346	3.300	3.258	3.227	3.184	3.156	3.122	3.102	3.054	3.039
	.005	4.442	4.192	4.026	3.909	3.822	3.754	3.700	3.656	3.620	3.563	3.511	3.473	3.421	3.388	3.346	3.322	3.265	3.246

Tabla A.10 (continuación)

m	α	υ = 10	12	14	16	18	20	22	24	26	30	35	40	50	60	80	100	250	500
9	.05	3.107	2.998	2.924	2.870	2.829	2.798	2.772	2.751	2.734	2.706	2.681	2.663	2.637	2.620	2.600	2.587	2.558	2.549
	.025	3.518	3.371	3.273	3.202	3.149	3.107	3.074	3.046	3.024	2.988	2.955	2.931	2.898	2.877	2.850	2.834	2.797	2.785
	.01	4.078	3.871	3.734	3.636	3.563	3.506	3.461	3.424	3.393	3.345	3.301	3.269	3.225	3.196	3.161	3.140	3.091	3.075
	.005	4.518	4.258	4.086	3.965	3.874	3.804	3.749	3.703	3.666	3.607	6.553	3.514	3.461	3.426	3.383	3.358	3.299	3.280
10	.05	3.169	3.055	2.977	2.921	2.878	2.845	2.819	2.797	2.779	2.750	2.724	2.704	2.678	2.660	2.639	2.626	2.596	2.586
	.025	3.581	3.428	3.326	3.252	3.197	3.153	3.119	3.091	3.067	3.030	2.996	2.971	2.937	2.915	2.887	2.871	2.832	2.820
	.01	4.144	3.930	3.787	3.686	3.610	3.552	3.505	3.467	3.435	3.385	3.340	3.307	3.261	3.232	3.195	3.174	3.123	3.107
	.005	4.587	4.318	4.140	4.015	3.922	3.850	3.792	3.745	3.707	3.646	3.591	3.551	3.496	3.460	3.416	3.391	3.330	3.310
11	.05	3.225	3.106	3.025	2.967	2.923	2.888	2.861	2.838	2.819	2.789	2.762	2.742	2.714	2.696	2.674	2.660	2.629	2.619
	.025	3.639	3.480	3.374	3.297	3.240	3.195	3.159	3.130	3.106	3.067	3.033	3.007	2.972	2.948	2.920	2.903	2.864	2.851
	.01	4.204	3.982	3.836	3.731	3.653	3.593	3.545	3.505	3.473	3.421	3.375	3.341	3.294	3.264	3.226	3.204	3.152	3.135
	.005	4.649	4.372	4.189	4.060	3.964	3.890	3.831	3.783	3.744	3.681	3.625	3.584	3.528	3.491	3.446	3.420	3.358	3.338
12	.05	3.277	3.153	3.069	3.008	2.963	2.927	2.899	2.875	2.856	2.825	2.797	2.776	2.747	2.729	2.705	2.692	2.659	2.649
	.025	3.691	3.527	3.417	3.339	3.279	3.233	3.196	3.166	3.141	3.102	3.066	3.039	3.003	2.979	2.950	2.933	2.892	2.879
	.01	4.259	4.031	3.880	3.773	3.692	3.630	3.581	3.540	3.507	3.454	3.407	3.372	3.324	3.293	3.254	3.232	3.179	3.161
	.005	4.706	4.422	4.234	4.102	4.004	3.928	3.867	3.818	3.777	3.714	3.656	3.614	3.556	3.519	3.473	3.446	3.383	3.362
13	.05	3.324	3.196	3.109	3.047	3.000	2.963	2.933	2.909	2.889	2.857	2.828	2.807	2.778	2.758	2.734	2.720	2.687	2.676
	.025	3.740	3.571	3.458	3.377	3.316	3.268	3.230	3.199	3.174	3.133	3.096	3.069	3.032	3.007	2.977	2.960	2.918	2.904
	.01	4.309	4.075	3.921	3.810	3.728	3.665	3.614	3.573	3.538	3.485	3.436	3.400	3.351	3.319	3.280	3.257	3.203	3.185
	.005	4.759	4.467	4.275	4.140	4.039	3.962	3.900	3.850	3.808	3.743	3.685	3.642	3.583	3.545	3.498	3.470	3.406	3.385
14	.05	3.368	3.236	3.146	3.082	3.034	2.996	2.965	2.941	2.920	2.887	2.857	2.836	2.805	2.785	2.761	2.747	2.712	2.701
	.025	3.785	3.611	3.495	3.412	3.349	3.301	3.262	3.230	3.204	3.162	3.124	3.096	3.058	3.033	3.003	2.984	2.942	2.928
	.01	4.357	4.117	3.958	3.846	3.762	3.697	3.645	3.603	3.567	3.513	3.463	3.426	3.376	3.344	3.304	3.280	3.225	3.207
	.005	4.809	4.510	4.314	4.175	4.073	3.994	3.931	3.879	3.837	3.771	3.711	3.667	3.607	3.568	3.521	3.493	3.427	3.406
15	.05	3.409	3.273	3.181	3.115	3.065	3.026	2.995	2.970	2.949	2.915	2.885	2.862	2.831	2.811	2.786	2.771	2.736	2.725
	.025	3.827	3.649	3.530	3.444	3.380	3.331	3.291	3.258	3.231	3.189	3.150	3.122	3.083	3.057	3.026	3.007	2.964	2.949
	.01	4.401	4.156	3.993	3.878	3.793	3.726	3.673	3.630	3.594	3.538	3.488	3.451	3.400	3.366	3.326	3.302	3.245	3.227
	.005	4.855	4.550	4.349	4.208	4.104	4.023	3.959	3.907	3.864	3.796	3.735	3.691	3.630	3.590	3.542	3.513	3.446	3.425
16	.05	3.448	3.308	3.214	3.146	3.095	3,055	3.023	2997	2.975	2.941	2.910	2.887	2.855	2.834	2.809	2.793	2.758	2.746
	.025	3.867	3.684	3.562	3.475	3.410	3.359	3.318	3.285	3.257	3.214	3.174	3.145	3.106	3.080	3.048	3.029	2.984	2.870
	.01	4.442	4.192	4.026	3.909	3.822	3.754	3.700	3.656	3.620	3.563	3.511	3.473	3.421	3.388	3.346	3.322	3.265	3.246
	.005	4.898	4.587	4.383	4.239	4.133	4.051	3.985	3.932	3.888	3.820	3.758	3.713	3.651	3.611	3.561	3.532	3.465	3.443

Fuente: Gill (1978).

Tabla A.11 Distribución del rango estudentizado para la DHS de Tukey ($\alpha = 0.05$)

Grados de libertad, ν	Número de tratamientos, k																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.0	27.0	32.8	37.2	40.5	43.1	45.4	47.3	49.1	50.6	51.9	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.09	8.33	9.80	10.89	11.73	12.43	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.36	16.57	16.77
3	4.50	5.91	6.83	7.51	8.04	8.47	8.85	9.18	9.46	9.72	9.95	10.16	10.35	10.52	10.69	10.84	10.98	11.12	11.24
4	3.93	5.04	5.76	6.29	6.71	7.06	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.67	8.80	8.92	9.03	9.14	9.24
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.42	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32	6.49	6.65	6.79	6.92	7.04	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.35	5.59	5.80	5.99	6.15	6.29	6.42	6.54	6.65	6.75	6.84	6.93	7.01	7.08	7.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.65
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.12	6.20	6.27	6.34	6.41	6.47
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.14	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.06	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.79	5.86	5.92	5.98	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.79	5.85	5.91	5.96
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.50	5.56	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.55	5.59
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.90	4.98	5.05	5.11	5.17	5.22	5.27	5.32	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.69	3.92	4.16	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4,74	4.80	4.84	4.89	4.93	4.97	5.01

Fuente: Snedecor y Cochran (1984).

Tabla A.12 Distribución del rango estudentizado de Duncan

gl \ p	α = 0.05								
	2	3	4	5	6	7	8	9	10
1	17.97								
2	6.09	6.09							
3	4.50	4.52	4.52						
4	3.93	4.01	4.03	4.03					
5	3.64	3.75	3.80	3.81	3.81				
6	3.46	3.59	3.65	3.68	3.69	3.70			
7	3.34	3.48	3.55	3.59	3.61	3.62	3.63		
8	3.26	3.40	3.48	3.52	3.55	3.57	3.57	3.58	
9	3.20	3.34	3.42	3.47	3.50	3.52	3.54	3.54	3.55
10	3.15	3.29	3.38	3.43	3.47	3.49	3.51	3.52	3.52
11	3.11	3.26	3.34	3.40	3.44	3.46	3.48	3.49	3.50
12	3.08	3.23	3.31	3.37	3.41	3.44	3.46	3.47	3.48
13	3.06	3.20	3.29	3.35	3.39	3.42	3.44	3.46	3.47
14	3.03	3.18	3.27	3.33	3.37	3.40	3.43	3.44	3.46
15	3.01	3.16	3.25	3.31	3.36	3.39	3.41	3.43	3.45
16	3.00	3.14	3.23	3.30	3.34	3.38	3.40	3.42	3.44
17	2.98	3.13	3.22	3.28	3.33	3.37	3.39	3.41	3.43
18	2.97	3.12	3.21	3.27	3.32	3.36	3.38	3.40	3.42
19	2.96	3.11	3.20	3.26	3.31	3.35	3.38	3.40	3.41
20	2.95	3.10	3.19	3.25	3.30	3.34	3.37	3.39	3.41
24	2.92	3.07	3.16	3.23	3.28	3.31	3.35	3.37	3.39
30	2.89	3.03	3.13	3.20	3.25	3.29	3.32	3.35	3.37
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35
60	2.83	2.98	3.07	3.14	3.20	3.24	3.28	3.31	3.33
120	2.80	2.95	3.04	3.12	3.17	3.22	3.25	3.29	3.31
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.27	3.29

gl \ p	α = 0.01								
	2	3	4	5	6	7	8	9	10
1	90.02								
2	14.04	14.04							
3	8.26	8.32	8.32						
4	6.51	6.68	6.74	6.76					
5	5.70	5.90	5.99	6.04	6.07				
6	5.24	5.44	5.55	5.62	5.66	5.68			
7	4.95	5.15	5.26	5.33	5.38	5.42	5.44		
8	4.74	4.94	5.06	5.13	5.19	5.23	5.26	5.28	
9	4.60	4.79	4.91	4.99	5.04	5.09	5.12	5.14	5.16
10	4.48	4.67	4.79	4.88	4.93	4.98	5.01	5.04	5.06
11	4.39	4.58	4.70	4.78	4.84	4.89	4.92	4.95	4.97
12	4.32	4.50	4.62	4.71	4.77	4.81	4.85	4.88	4.91
13	4.26	4.44	4.56	4.64	4.71	4.75	4.79	4.82	4.85
14	4.21	4.39	4.51	4.59	4.66	4.70	4.74	4.77	4.80
15	4.17	4.34	4.46	4.55	4.61	4.66	4.70	4.73	4.76
16	4.13	4.31	4.43	4.51	4.57	4.62	4.66	4.70	4.72
17	4.10	4.27	4.39	4.47	4.54	4.59	4.63	4.66	4.69
18	4.07	4.25	4.36	4.45	4.51	4.56	4.60	4.64	4.66
19	4.05	4.22	4.33	4.42	4.48	4.53	4.57	4.61	4.64
20	4.02	4.20	4.31	4.40	4.46	4.51	4.55	4.59	4.62
24	3.96	4.13	4.24	4.32	4.39	4.44	4.48	4.52	4.55
30	3.89	4.06	4.17	4.25	4.31	4.36	4.41	4.45	4.48
40	3.82	3.99	4.10	4.18	4.24	4.29	4.33	4.38	4.41
60	3.76	3.92	4.03	4.11	4.18	4.23	4.27	4.31	4.34
120	3.70	3.86	3.97	4.04	4.11	4.16	4.20	4.24	4.27
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.13	4.17	4.21

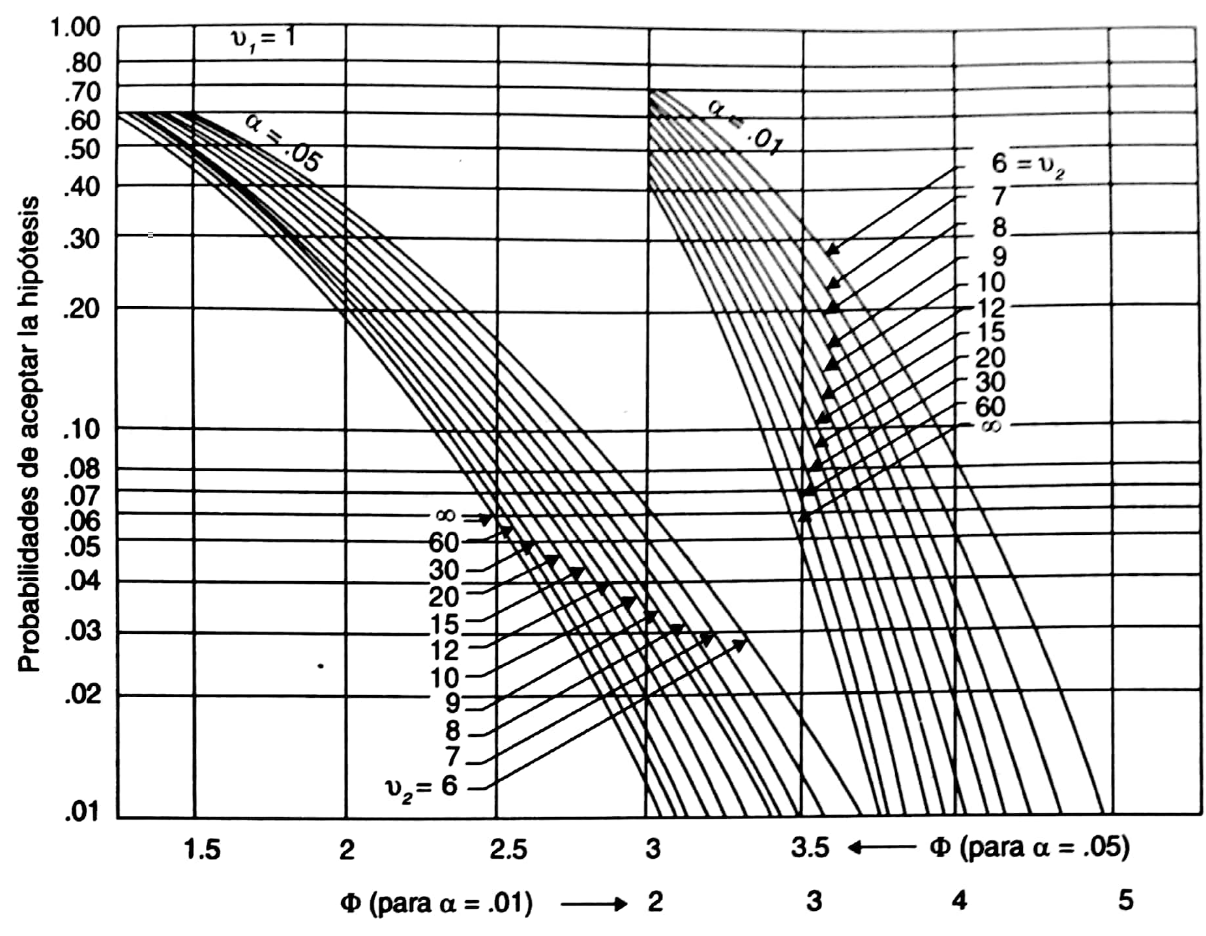
Fuente: Miller y Freund (1965).

Tabla A.13 Valores para t_D de Dunnett (prueba bilateral)

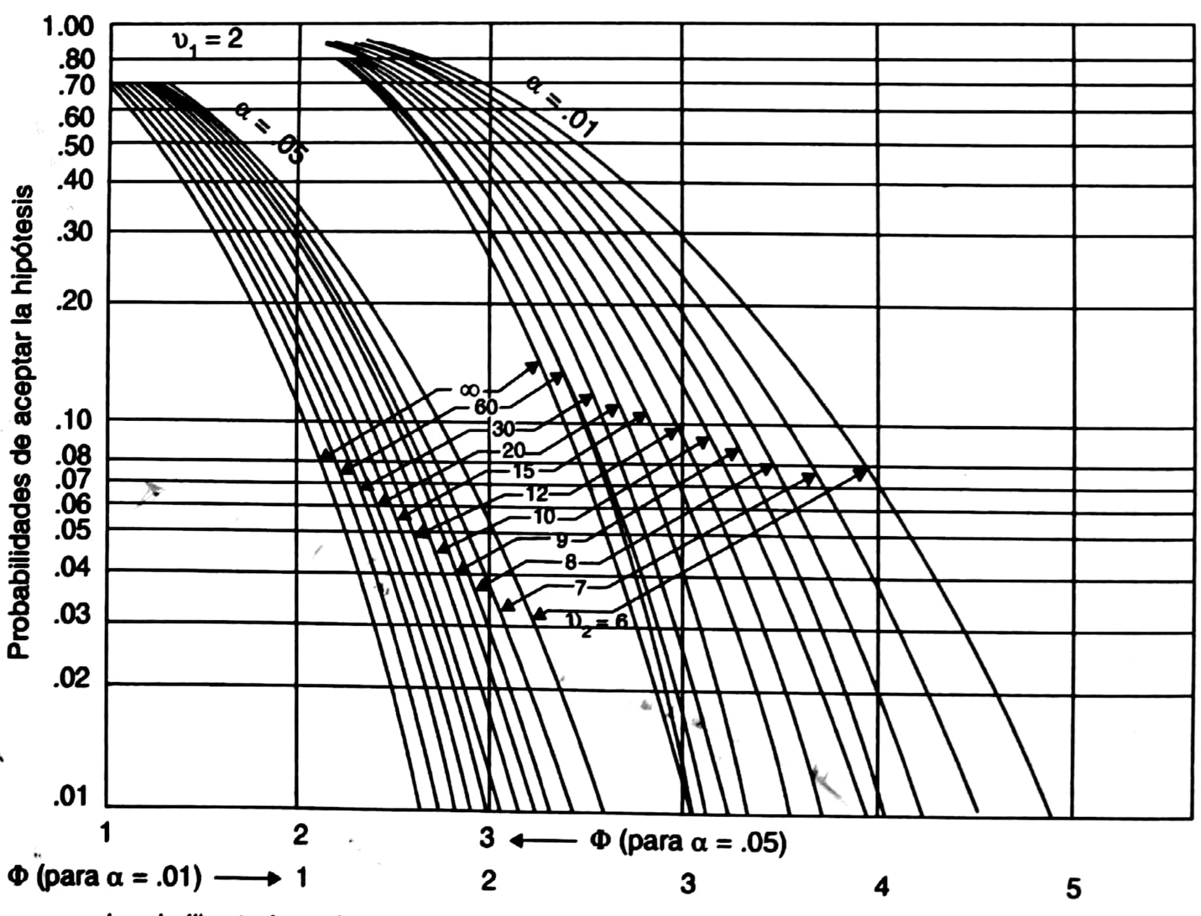
		k – 1						
υ	α	*1*	*2*	*3*	*4*	*5*	*6*	*7*
5	.05	2.57	3.03(2.3)	3.29(3.6)	3.48(4.6)	3.62(5.4)	3.73(5.9)	3.82(6.4)
	.01	4.03	4.63(1.8)	4.98(3.0)	5.22(3.9)	5.41(4.6)	5.56(5.2)	5.69(5.7)
6	.05	2.45	2.86(2.1)	3.10(3.4)	3.26(4.3)	3.39(5.0)	3.49(5.6)	3.57(6.0)
	.01	3.71	4.21(1.6)	4.51(2.7)	4.71(3.5)	4.87(4.1)	5.00(4.6)	5.10(5.1)
7	.05	2.36	2.75(2.0)	2.97(3.2)	3.12(4.1)	3.24(4.8)	3.33(5.3)	3.41(5.7)
	.01	3.50	3.95(1.5)	4.21(2.4)	4.39(3.1)	4.53(3.7)	4.64(4.2)	4.74(4.6)
8	.05	2.31	2.67(2.0)	2.88(3.1)	3.02(3.9)	3.13(4.5)	3.22(5.1)	3.29(5.5)
	.01	3.36	3.77(1.3)	4.00(2.2)	4.17(2.9)	4.29(3.4)	4.40(3.9)	4.48(4.2)
9	.05	2.26	2.61(1.9)	2.81(3.0)	2.95(3.8)	3.05(4.4)	3.14(4.9)	3.20(5.3)
	.01	3.25	3.63(1.2)	3.85(2.1)	4.01(2.7)	4.12(3.2)	4.22(3.6)	4.30(3.9)
10	.05	2.23	2.57(1.8)	2.76(2.9)	2.89(3.6)	2.99(4.2)	3.07(4.7)	3.14(5.1)
	.01	-3.17	3.53(1.2)	3.74(1.9)	3.88(2.5)	3.99(3.0)	4.08(3.4)	4.16(3.7)
11	.05	2.20	2.53(1.8)	2.72(2.8)	2.84(3.5)	2.94(4.1)	3.02(4.6)	3.08(4.9)
	.01	3.11	3.45(1.1)	3.65(1.8)	3.79(2.4)	3.89(2.8)	3.98(3.2)	4.05(3.5)
12	.05	2.18	2.50(1.7)	2.68(2.7)	2.81(3.4)	2.90(4.0)	2.98(4.4)	3.04(4.8)
	.01	3.05	3.39(1.1)	3.58(1.7)	3.71(2.3)	3.81(2.7)	3.89(3.0)	3.96(3.3)
13	.05	2.16	2.48(1.7)	2.65(2.7)	2.78(3.4)	2.87(3.9)	2.94(4.3)	3.00(4.7)
	.01	3.01	3.33(1.0)	3.52(1.7)	3.65(2.2)	3.74(2.6)	3.82(2.9)	3.89(3.2)
14	.05	2.14	2.46(1.7)	2.63(2.6)	2.75(3.3)	2.84(3.8)	2.91(4.2)	2.97(4.6)
	.01	2.98	3.29(1.0)	3.47(1.6)	3.59(2.1)	3.69(2.5)	3.76(2.8)	3.83(3.0)
15	.05	2.13	2.44(1.7)	2.61(2.6)	2.73(3.2)	2.82(3.8)	2.89(4.2)	2.95(4.5)
	.01	2.95	3.25(0.9)	3.43(1.5)	3.55(2.0)	3.64(2.4)	3.71(2.7)	3.78(2.9)
16	.05	2.12	2.42(1.6)	2.59(2.5)	2.71(3.2)	2.80(3.7)	2.87(4.1)	2.92(4.4)
	.01	2.92	3.22(0.9)	3.39(1.5)	3.51(1.9)	3.60(2.3)	3.67(2.6)	3.73(2.8)
17	.05	2.11	2.41(1.6)	2.58(2.5)	2.69(3.1)	2.78(3.6)	2.85(4.0)	2.90(4.4)
	.01	2.90	3.19(0.9)	3.36(1.5)	3.47(1.9)	3.56(2.2)	3.63(2.5)	3.69(2.7)
18	.05	2.10	2.40(1.6)	2.56(2.5)	2.68(3.1)	2.76(3.6)	2.83(4.0)	2.89(4.3)
	.01	2.88	3.17(0.9)	3.33(1.4)	3.44(1.8)	3.53(2.2)	3.60(2.4)	3.66(2.7)
19	.05	2,09	2.39(1.6)	2.55(2.5)	2.66(3.1)	2.75(3.5)	2.81(3.9)	2.87(4.2)
	.01	2.86	3.15(0.9)	3.31(1.4)	3.42(1.8)	3.50(2.1)	3.57(2.4)	3.63(2.6)
20	.05	2.09	2.38(1.6)	2.54(2.4)	2.65(3.0)	2.73(3.5)	2.80(3.9)	2.86(4.2)
	.01	2.85	3.13(0.8)	3.29(1.4)	3.40(1.7)	3.48(2.1)	3.55(2.3)	3.60(2.5)
24	.05	2.06	2.35(1.5)	2.51(2.3)	2.61(2.9)	2.70(3.4)	2.76(3.7)	2.81(4.0)
	.01	2.80	3.07(0.8)	3.22(1.3)	3.32(1.6)	3.40(1.9)	3.47(2.1)	3.52(2.4)
30	.05	2.04	2.32(1.5)	2.47(2.3)	2.58(2.8)	2.66(3.2)	2.72(3.6)	2.77(3.9)
	.01	2.75	3.01(0.7)	3.15(1.2)	3.25(1.5)	3.33(1.8)	3.39(2.0)	3.44(2.2)
40	.05	2.02	2.29(1.4)	2.44(2.2)	2.54(2.7)	2.62(3.1)	2.68(3.4)	2.73(3.7)
	.01	2.70	2.95(0.7)	3.09(1.1)	3.19(1.4)	3.26(1.6)	3.32(1.8)	3.37(2.0)
60	.05	2.00	2.27(1.4)	2.41(2.1)	2.51(2.6)	2.58(3.0)	2.64(3.3)	2.69(3.5)
	.01	2.66	2.90(0.6)	3.03(1.0)	3.12(1.3)	3.19(1.5)	3.25(1.6)	3.29(1.8)
120	.05	1.98	2.24(1.3)	2.38(2.0)	2.47(2.5)	2.55(2.8)	2.60(3.1)	2.65(3.3)
	.01	2.62	2.85(0.6)	2.97(0.9)	3.06(1.1)	3.12(1.3)	3.18(1.5)	3.22(1.6)
∞	.05	1.96	2.21(1.3)	2.35(1.9)	2.44(2.3)	2.51(2.7)	2.57(2.9)	2.61(3.1)
	.01	2.58	2.79(0.5)	2.92(0.8)	3.00(1.0)	3.06(1.2)	3.11(1.3)	3.15(1.4)

Fuente: **Gill (1978).**

B. Curvas características de operación para el modelo de efectos fijos en el análisis de varianza



υ_1 = grados de libertad en el numerador, υ_2 = grados de libertad en el denominador



υ_1 = grados de libertad en el numerador, υ_2 = grados de libertad en el denominador

$\upsilon_1 = 3$

$\alpha = .05$

$\alpha = .01$

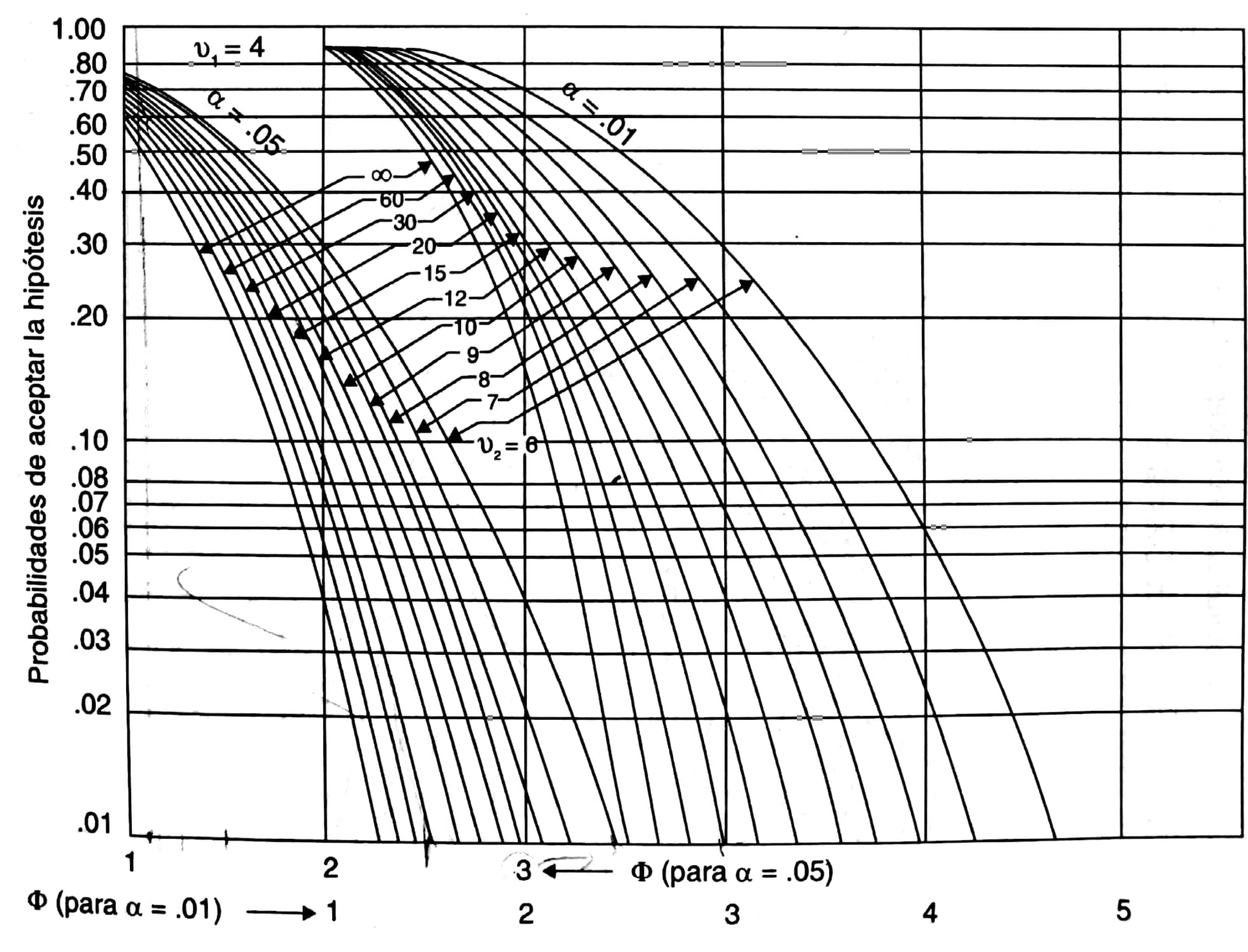
∞
60
30
20
15
12
10
9
8
7
$\upsilon_2 = 6$

Probabilidades de aceptar la hipótesis

1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01

1
2
3 ← Φ (para α = .05)

Φ (para α = .01) → 1
2
3
4
5



$\upsilon_1 = 5$

$\alpha = .05$ $\alpha = .01$

Probabilidades de aceptar la hipótesis

1.00 .80 .70 .60 .50 .40 .30 .20 .10 .08 .07 .06 .05 .04 .03 .02 .01

$6 = \upsilon_2$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

∞, 60, 30, 20, 15, 12, 10, 9, 8, 7, $\upsilon_2 = 6$

1 2 3 ← Φ (para α = .05)

Φ (para α = .01) → 1 2 3 4 5

$\upsilon_1 = 6$

$\alpha = .05$ $\alpha = .01$

Probabilidades de aceptar la hipótesis

1.00 .80 .70 .60 .50 .40 .30 .20 .10 .08 .07 .06 .05 .04 .03 .02 .01

$6 = \upsilon_2$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

$\upsilon_2 = 6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

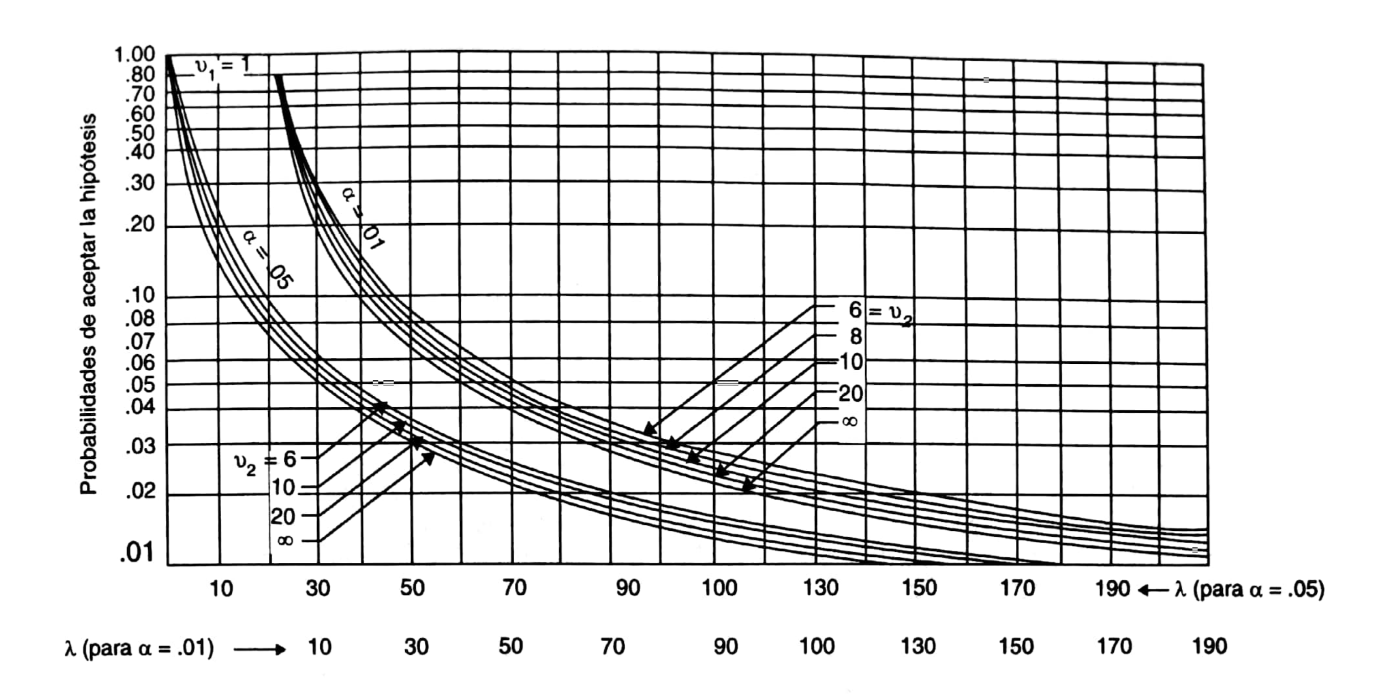
1 2 3 ← Φ (para α = .05)

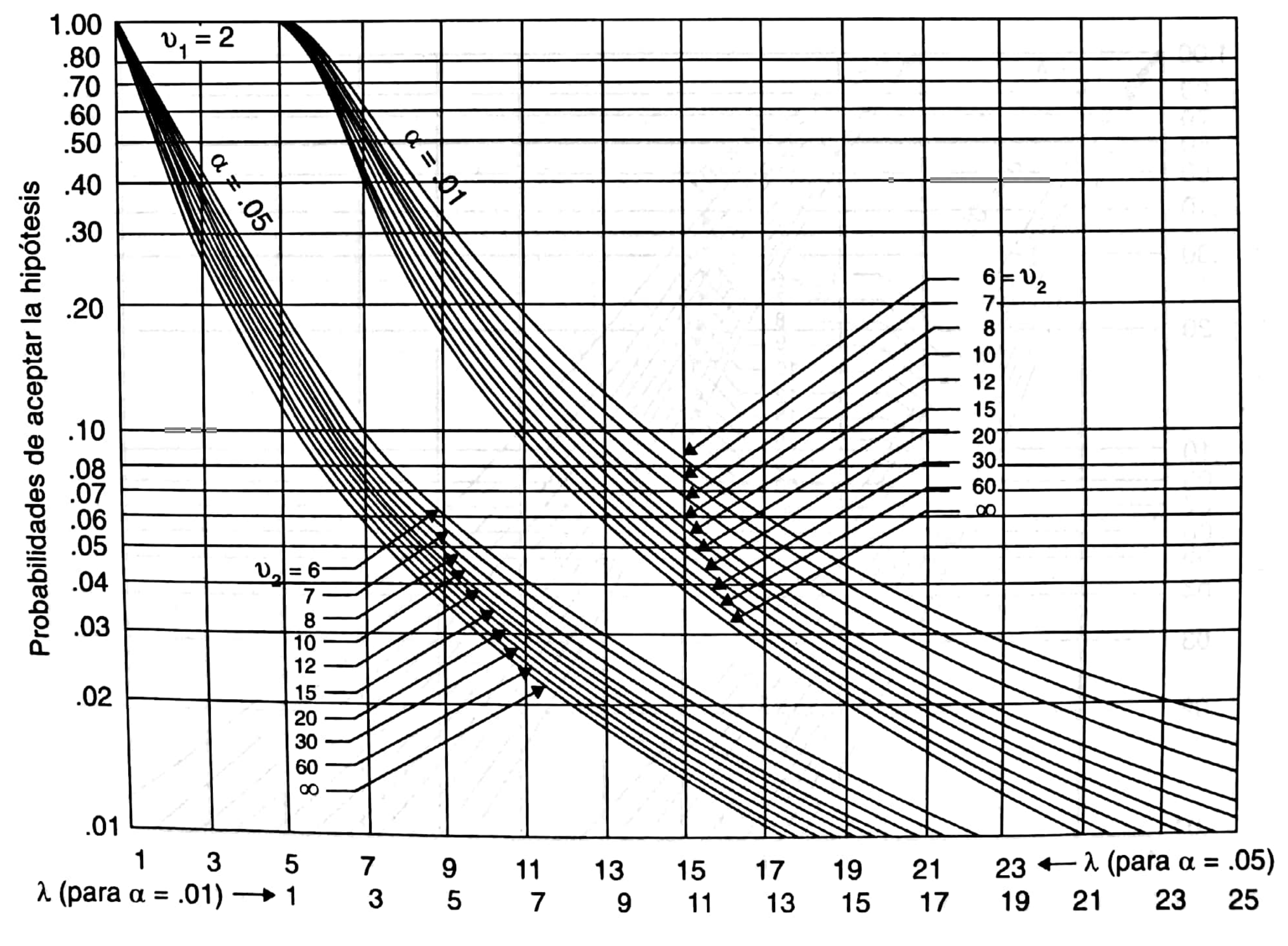
Φ (para α = .01) → 1 2 3 4

$\upsilon_1 = 7$
$\alpha = .05$
$\alpha = .01$
Probabilidades de aceptar la hipótesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
$6 = \upsilon_2$
7
8
9
10
12
15
20
30
60
∞
$\upsilon_2 = 6$
7
8
9
10
12
15
20
30
60
∞
1
2
3 ← Φ (para α = .05)
Φ (para α = .01) → 1
2
3
4

$\upsilon_1 = 8$
$\alpha = .05$
$\alpha = .01$
Probabilidades de aceptar la hipótesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
$6 = \upsilon_2$
7
8
9
10
12
15
20
30
60
∞
$\upsilon_2 = 6$
7
8
9
10
12
15
20
30
60
∞
1
2
3 ← Φ (para α = .05)
Φ (para α = .01) → 1
2
3
4

C. Curvas características de operación para el modelo de efectos aleatorios en el análisis de varianza


Probabilidades de aceptar la hipótesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
υ1 = 1
α = .05
α = .01
υ2 = 6
10
20
∞
6 = υ2
8
10
20
∞
10 30 50 70 90 100 130 150 170 190 ← λ (para α = .05)
λ (para α = .01) → 10 30 50 70 90 100 130 150 170 190


Probabilidades de aceptar la hipótesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
υ1 = 2
α = .05
α = .01
υ2 = 6
7
8
10
12
15
20
30
60
∞
6 = υ2
7
8
10
12
15
20
30
60
∞
1 3 5 7 9 11 13 15 17 19 21 23 ← λ (para α = .05)
λ (para α = .01) → 1 3 5 7 9 11 13 15 17 19 21 23 25

Probabilidades de aceptar la hipótesis

$\upsilon_1 = 3$

$\alpha = .05$

$\alpha = .01$

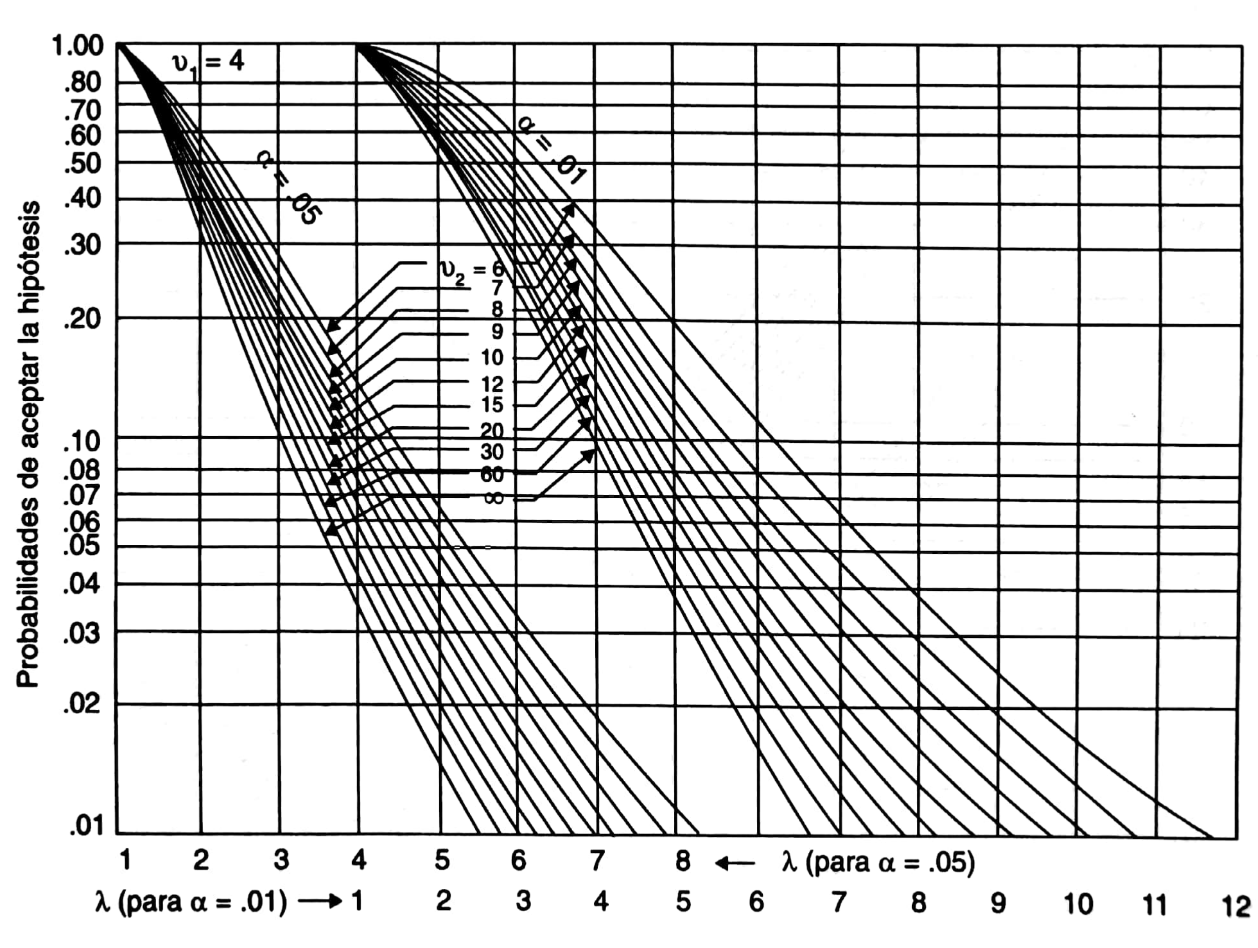
$\upsilon_2 = 6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

$6 = \upsilon_2$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

1.00 .80 .70 .60 .50 .40 .30 .20 .10 .08 .07 .06 .05 .04 .03 .02 .01

1 2 3 4 5 6 7 8 9 10 11 ← λ (para α = .05)

λ (para α = .01) ⟶ 1 2 3 4 5 6 7 8 9 10 11 12 13



$\upsilon_1 = 5$

$\alpha = .05$ $\alpha = .01$

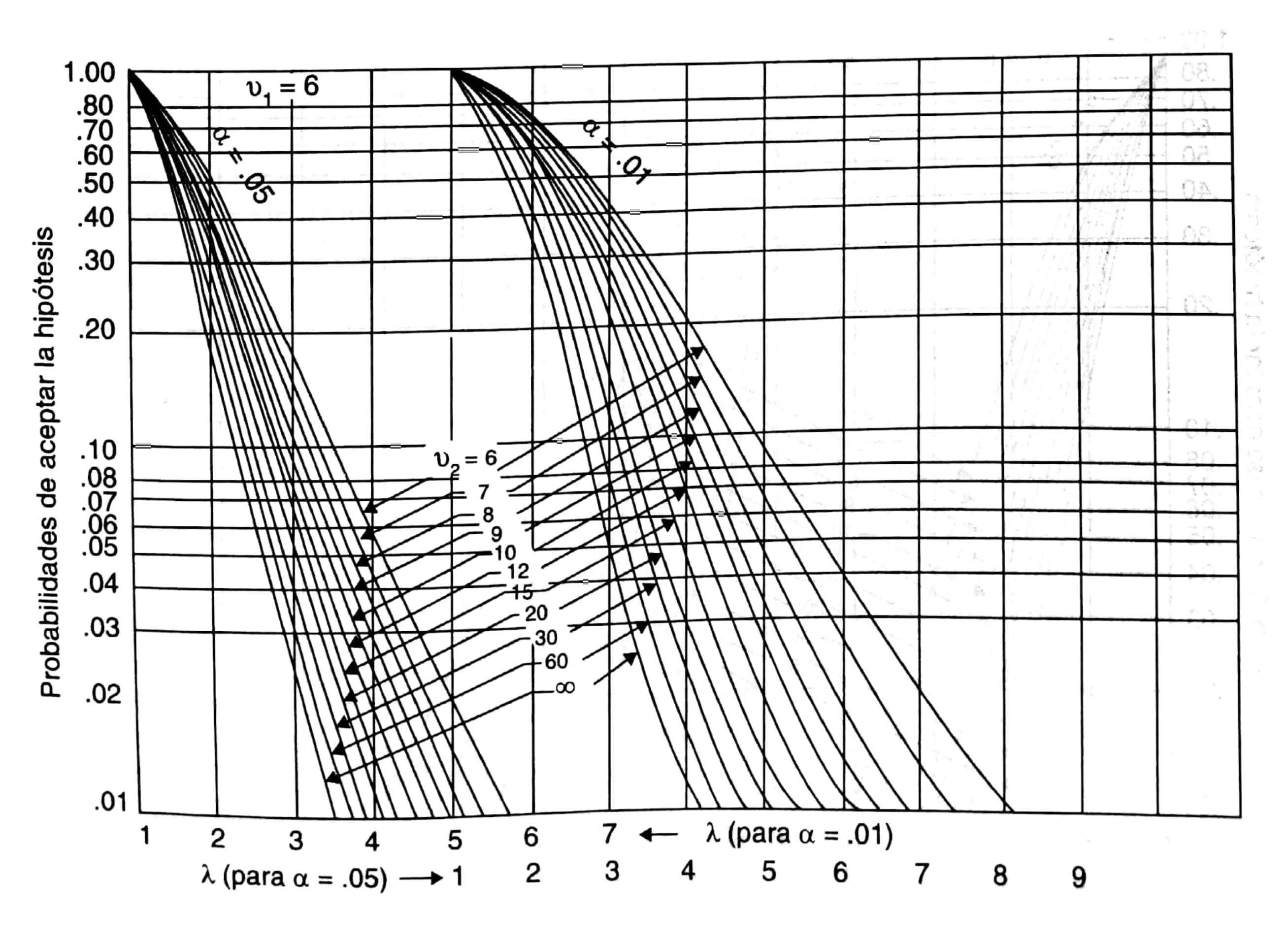
$\upsilon_2 = 6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

Probabilidades de aceptar la hipótesis

1.00 .80 .70 .60 .50 .40 .30 .20 .10 .08 .07 .06 .05 .04 .03 .02 .01

1 2 3 4 5 6 7 ← λ (para α = .05)

λ (para α = .01) → 1 2 3 4 5 6 7 8 9 10



$\upsilon_1 = 7$

$\alpha = .05$ $\alpha = .01$

Probabilidades de aceptar la hipótesis

1.00 .80 .70 .60 .50 .40 .30 .20 .10 .08 .07 .06 .05 .04 .03 .02 .01

$\upsilon_2 = 6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

1 2 3 4 5 6 ← λ (para $\alpha = .05$)

λ (para $\alpha = .01$) → 1 2 3 4 5 6 7 8

$\upsilon_1 = 8$

$\alpha = .05$ $\alpha = .01$

Probabilidades de aceptar la hipótesis

1.00 .80 .70 .60 .50 .40 .30 .20 .10 .08 .07 .06 .05 .04 .03 .02 .01

$\upsilon_2 = 6$, 7, 8, 9, 10, 12, 15, 20, 30, 60, ∞

1 2 3 4 5 6 ← λ (para $\alpha = .01$)

λ (para $\alpha = .05$) → 1 2 3 4 5 6 7 8

Índice analítico

F

G

H

I

L

M

N

O

P

R

S